Journal of Development Economics 171 (2024) 103330

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

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

Abhijit V. Banerjee[a], Gaurav Chiplunkar[b,*]

[a] *Massachusetts Institute of Technology, Room 540, Morris and Sophie Chang Building 50, Memorial Drive, Cambridge, 02142, MA, USA*
[b] *Darden Business School, University of Virginia, 100, Darden Blvd., Charlottesville, 22902, VA, USA*

ARTICLE INFO

MSC:
J22
J24
J61
J62
J64

Keywords:
Matching
Labor
Job-search
Recruitment
Experiment

ABSTRACT

This paper provides evidence of matching frictions in the Indian labor market. Using several methods to elicit genuine preferences of job-seekers over jobs, we show that: (a) there is substantial variation in job-seekers preferences over the same jobs; and (b) placement officers, responsible for placing job-seekers in jobs, have poor knowledge of it. Providing placement offers with this information improves matching of job-seekers to interviews, even after taking into account redistribution of jobs across job-seekers. Treated job-seekers get more preferred jobs and retain them in the short run (three months), but not in the longer run (six months).

1. Introduction

Youth unemployment remains a major policy challenge facing the world today. For example, using data from the latest Labor Force Surveys across 150 countries,[1] younger workers (ages 20–24) are on average three-four times more likely to be available and seeking work as compared to older ones (see Fig. 1).[2] In the Indian context as well, the empirical focus for our study, the unemployment rate is 21% for workers between the age of 20–24, 10% between the age of 25–29, and less than 1% by the age of 40–44. Search and matching frictions in the labor market have been increasingly advocated as an important channel to explain why it takes a long time for these young job-seekers to find the job they want. For example, thick market externalities, or distortions in the job search process make it possible that a job seeker searches too little.[3] 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 across the world[4] and evaluating their impact has been the focus of a growing and recent literature.[5]

This paper reports on a randomized trial that examines a friction that has been relatively understudied: the implications of inefficient

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

* Corresponding author.
E-mail addresses: banerjee@mit.edu (A.V. Banerjee), chiplunkarg@darden.virginia.edu (G. Chiplunkar).

1 Source: ILO Data Explorer.

2 As reported in Figure A1, this pattern is very robust across both genders as well as using an alternate age group of 25–29 years instead of 20–24 years.

3 See Diamond (1982), Mortensen and Pissarides (1994) and Acemoglu (1996, 1997).

4 See Card et al. (2010, 2018), Escudero et al. (2019), McKenzie (2017) for a meta-analyses of studies.

5 See for example studies by Dammert et al. (2015), Altmann et al. (2018) and Belot et al. (2019), who provide information on vacancies to job-seekers; Coles et al. (2010, 2013), who study labor markets signals in the job market for economists; Beam (2016) and Abebe et al. (2023), who test the impact of job fairs; Banerjee and Sequeira (2023), Abebe et al. (2021) and Franklin (2018), who subsidize job search; Abel et al. (2020), Alfonsi et al. (2020), Groh et al. (2015), Bassi and Nansamba (2022) and Pallais (2014), who reduce screening costs through reference letters, skill report cards, vocational training, and referrals.

https://doi.org/10.1016/j.jdeveco.2024.103330
Received 21 March 2023; Received in revised form 8 May 2024; Accepted 7 June 2024
Available online 12 June 2024
0304-3878/

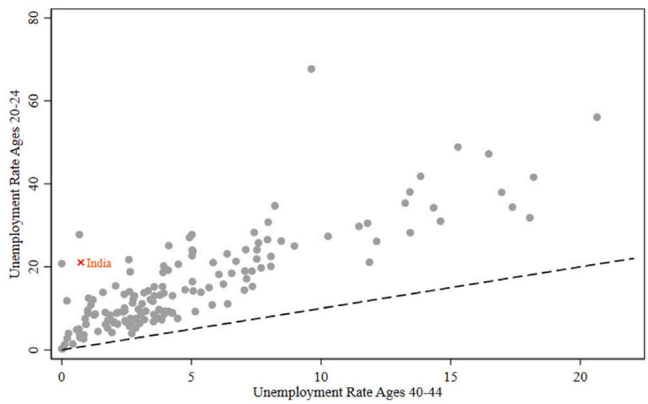

Fig. 1. Youth unemployment rates across countries. *Notes:* The above graph uses data from the latest round of labor force surveys from the International Labor Organization (ILO) for 150 countries between 2015–2019. Unemployment Rates are reported for ages 20–24 and 40–44, which are the share of the labor force in that age group that is without work but available for and seeking employment.

matching of job-seekers to jobs by intermediaries in the labor market. Intermediaries are widely prevalent in labor markets across the world. In a survey of over 4000 Human Resource departments across 35 countries, one in three firms reported routinely hiring workers through intermediaries such as third-party recruiters and staffing firms. Similarly, these intermediaries were also a common way through which job-seekers heard about jobs, second only to referrals (LinkedIn, 2015, 2017). In India as well, firms have been increasingly hiring workers "off-the-books" on temporary contracts to circumvent cumbersome labor laws (Chaurey, 2015; Chiplunkar et al., 2023). These workers are usually hired through third-party intermediaries such as external consultants, agency firms and educational/training institutes (India Skills Report, 2019), and account for over a third of total employment in formal jobs in India (Bertrand et al., 2021). Lastly, the low/medium-skilled Indian labor market, similar to other developing countries, is characterized by seasonal employment, high turnover rates and low-retention in jobs. This makes it costly for firms to find, train and trust workers, further amplifying the role of intermediaries, and why workers and firms continue to rely on them (Mamgain, 2019). Put together, the above indicates that poor matching of job-seekers to jobs by labor market intermediaries could be one potential explanation for high youth unemployment rates in a developing economy like India, where there seems to be no dearth of these jobs per se.

Our empirical setting studies the role of placement managers in impacting the employment outcomes of young trainees after the completion of a vocational training program. In partnership with a large vocational training institute, we first start by providing detailed evidence that these placement managers often have little information about the job preferences of the trainees that they are responsible for placing. As a result, managers often offer these trainees interviews for jobs that they have no interest in. However, to document this mismatch, one would need to reliably know the preferences of these job-seekers. Otherwise, what we may believe to be a mismatch, could in fact reflect managers knowing more about job-seeker preferences than we do. Unfortunately, getting people to reliably reveal their preferences is not easy, especially when preferences are multi-dimensional so that standard a Becker–DeGroot–Marschak (BDM) mechanism cannot be used. To elicit preferences, we asked these potential job-seekers, currently trainees at the vocational institute, to rank a list of real-world job options. These jobs were carefully chosen to resemble their future jobs, with variation along specific job characteristics. To test whether these rankings reflect their true underlying preferences, for half the job-seekers, we amplified their incentives to be strategic in ranking these jobs, while not doing so for the remaining (see Section 3.2). The two preference distributions we get are essentially identical, giving us some confidence that (a) we do not need strong incentives to elicit true preferences and (b) these are their actual preferences rather than what they would report strategically to maximize their chances of getting a job.

With the (true) preferences of these job-seekers, we then ask their placement managers to predict the preferences for each trainee over the *same* set of jobs used for the elicitation exercise described above. Specifically, we ask the manager to pick the three best jobs (in order of preference) from that trainee's point of view. Through various measures on how the manager's ranking correlates with the trainee's ranking, we show that managers do better than just picking jobs at random, but are far from knowing trainee preferences perfectly. For example, the job picked by the manager as the best job for a trainee is ranked on average at 7.2 by the trainee herself on a scale of 1 to 11 (1 is the worst and 11 is the best). If the manager had picked at random instead, the average rank would have been 5.5 and if the manager knew the preference perfectly, the rank should have been 11. Furthermore, we do not find any evidence that managers are systematically better at predicting preferences for some trainees more than others. In fact, a rich set of trainee, job, manager, trade and center characteristics (selected using machine learning methods) is only able to explain 13% of the variation in trainee preferences (see Section 4.2). Lastly, we explore whether managers systematically lack knowledge along specific job dimensions (such as salary, location, etc.). We find that managers underestimate the willingness of trainees to migrate to cities for jobs. They therefore provide trainees with interviews for lower paying, local jobs instead of the higher paying ones in the cities (see Section 4.3).

Having documented managers' lack of knowledge of trainee preferences, the second part of the study (Section 5 onward), discusses the implementation and impact of a randomized control trial. We experimentally vary the information that placement managers have about trainee preferences. For half of the trainees in a batch (henceforth the Treatment group), we provide the manager with details for the four most preferred jobs for each trainee, and not for the others (Control group). We show that this intervention substantially improves the allocation of interviews, as trainees in the Treatment group are more likely to get an interview for their most preferred jobs (see Table A11). However, though trainees in the Treatment group benefit from the information being provided to the manager, it does not necessarily mean that the overall matching becomes more efficient, since there could be displacement effects (Crépon et al., 2013; Cheung et al., 2019). To get at this, in an ideal experiment, one would randomize information on trainee preferences at the batch level instead of across trainees within a batch (as we do). Our choice of the latter, as discussed in Section 6, was primarily driven by limited resources and logistical challenges of operating in the rural areas of Uttar Pradesh (such as the uncertainty of completing batches, challenges of surveying in remote areas, etc.). Therefore, the limitation of our design (that could be overcome by randomizing at the batch level) is that we do not have a counterfactual allocation of interviews within the batch in the absence of our intervention.

We take two approaches to overcome this limitation (see Section 6). First, we hypothesize that if managers shrink their assessments of individual-specific preferences towards the group means, it will likely induce congestion in job allocation. Our intervention, by providing these job preferences for a subset of individuals, can therefore help reduce this congestion and improve overall matching efficiency without significant spillovers. A key assumption under this hypothesis that we empirically test and confirm (see Section 6.1), is that the dispersion in managers' assessment of trainee preferences for a job j is indeed lower than the (true) underlying variation in trainee preferences i.e., managers do "shrink" their assessment of trainee preferences. Our second approach is more theoretical in nature. Note that our primary challenge is to reasonably predict the allocation of interviews across trainees in the absence of our intervention. To generate this counterfactual allocation, we first model the decision-making rule of a placement manager. We make three possible assumptions about what a manager knows–(i) a *complete information* case, where she knows what we know about trainee preferences; (ii) a *no information* case, where she knows what she tells us in our manager interviews about trainee preferences; (iii) a *hybrid information* case, where she knows what we tell her for the Treatment group but what she tells us for the Control group. Under these alternative assumptions, we ask whether a stable matching algorithm can predict the allocation we see in the data. We find (not surprisingly) that the complete information case does poorly at explaining the empirical allocation of interviews, and the hybrid information case fits the data the best. In other words, the manager does come close to achieving efficiency, subject to her information constraints. Given this, we can then simulate a counterfactual of how interviews would have been allocated in the absence of our intervention ("no information" case described above), and compare it to when we provide her with preferences for some trainees ("hybrid information" case). We find that trainees in the Treatment group are 8–11 pp more likely to get an interview for at least one of their (up to) four most preferred jobs, while those in the Control group remain unaffected on average (see Table 4). At least by this metric, the intervention was a success.

The final section of the paper (see Section 7) asks whether the success in altering the allocation of interviews has differential employment consequences for the trainees. In particular, do they get jobs that they like better, and does that experience make them more likely to stay employed? The results suggest that treated trainees are no more likely to get interviews and offers for jobs, nor are they more likely to accept them, relative to their Control counterparts. This is consistent with placement managers trying to be fair to all students. However treated trainees do get interviews and offers for jobs that they like better, and are more likely to accept them as well. Given that this is a high turnover environment,[6] we then examine whether being matched to preferred jobs increases job retention and labor market adherence. The evidence here is a bit more mixed. We find that treated trainees are 1/3 more likely to be in a job they were placed in 3 months after placement, and 1/6th more likely to be employed anywhere (Panel B, Table 5). After six months, almost no one is employed in their original jobs,[7] but conditional on accepting a job after training, treated trainees are 70% more likely to be employed anywhere (p=0.09).

Finally, we examine the impact of our intervention on job retention and job quality by calculating a preference-weighted measure of job outcomes (denoted by y_i^{pw}). y_i^{pw} aggregates outcomes across jobs after weighting them by individuals' preference for it. We normalize y_i^{pw} to have mean zero and standard deviation 1 for the Control group to make the interpretation of comparisons easier. From Columns 5–6 of Table 5, Treatment trainees have 0.27σ better quality of interviews,[8] 0.19σ better quality offers, 0.24σ higher acceptance of these offers, and are 0.25σ more likely to stay them three months after placement as well. In fact, our job ranking exercise makes a broader point: we show that getting better (more preferred) interviews, irrespective of a trainee's treatment status, do translate into better placement and employment outcomes for trainees more generally. Our results therefore suggest that the initial matching frictions are substantial (around 20%–25% of a trainee's average monthly salary) and using low-cost methods to mitigate them does improve the quality of placements. However, these matches do not seem to persist over the longer run a feature that is documented in other low-income labor markets as well. While we provide some reasons to suggest that trainees appear to update their beliefs and preferences after being explosed to certain job environments, labor markets, and urban living conditions, understanding the drivers (aspirations, norms, etc.) behind why young job-seekers make these employment choices seems key to improving the targeting of policy interventions related to employment and job search.

One might wonder why placement managers do not exert effort themselves to elicit trainee preferences. Qualitative interviews with these managers provide multiple reasons: first, given the high attrition in jobs, managers are usually interested in understanding whether a trainee wants to work or not, as opposed where and for how long. Second, conditional on wanting to work, managers do informally solicit preferences, but at a relatively aggregate level rather than for each trainee. As we know from study, there is in fact a substantial heterogeneity in preferences across trainees, so this is a crucial distinction. Third, a part of the answer also seems to be that since placement managers also manage the day-to-day activities of the center (such as publicizing and starting new batches, enrollments, staff payments, etc.), they are constrained in their time. In fact only 2 out of 10 managers in our survey said they would "like to do more". The other 8 said they were taking adequate efforts, or in fact, doing much more than needed. All of them reported not finding enough time to talk to each trainee about his/her preferences or to put it in other words, it was 'costly' for them to acquire this information. It is plausible that perhaps they overestimate just how difficult it is to learn something useful. In fact, the literature on the internal management of firms in developing countries explores this very question and finds that managers in even larger firms do not always take what appear to be quite easy steps to enhance their productivity, but switch to doing so when they are suggested by outside consultants (Bloom and Van Reenen, 2007, 2010; Bloom et al., 2013). We see our managers partly through the same lens.

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

2. Context and data

2.1. Institutional setting

Despite high unemployment among the Indian youth (discussed earlier), a widely cited survey on 'Labor/Skill Shortage for Industry' conducted by the Federation of Indian Chambers of Commerce and Industry[9] found that 90% of firms reported facing shortage of labor and 89% reported a shortage of labor. This indicates (among other things) a potential mismatch between labor demand and supply and it is therefore not surprising that active labor market policies have been at the center of Indian policy over 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. An ambitious targeting of training

[6] In our sample, 48 (72)% and 10 (38)% of Control trainees are employed in the same (any) job after three and six months conditional on accepting a job after training.

[7] These characteristics seem to be a consistent feature across other similar labor markets, such as those in Ethiopia (Abebe et al., 2023; Blattman et al., 2019), Uganda (Alfonsi et al., 2020; Bassi and Nansamba, 2022) and the Philippines (Beam, 2016). This could be because of seasonality in agriculture, adverse events at home, marriage and fertility decisions, or just challenges of living alone in a city.

[8] A quick back-of-the-envelope calculation suggests that a 0.1σ higher job quality is equivalent to 8%–10% higher monthly salary.

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

over 500 million people by 2022 was set through public–private partnerships that would be managed by the National Skill Development Corporation (NSDC). While the NSDC designed the components of training programs under the Skill India Mission, the private sector was incentivized to undertake their implementation through financial payouts after the successful completion of a training program. An important aspect of this compensation was that 15%–20% it (for the short training courses) was contingent on trainees being placed and employed for three months after the completion of the training program.

On the impact of training programs in India, a study conducted by the International Labour Organization (2003) in Andhra Pradesh, Maharashtra and Odisha found poor labor market outcomes for the trainees after the training program. Another study by the World Bank (2008) found that a high proportion of trainees remain unemployed after the training program. Furthermore, more recent reports from the impact of training programs (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 what these programs deliver, and what their clients want. This could be either a lack of jobs that the clients want, or because the existing pool of jobs are not allocated to the right set of applicants.

2.2. Study context

Sample description: For this study, we partner with Skills Academy, a large training institute that undertakes the design, management and implementation of training programs across 17 states in India.[10] 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. Our study sample consists of 538 trainees who are enrolled in 28 field-specific training programs or "batches" across 10 centers in the states of Uttar Pradesh and the National Capital 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 trainees are enrolled in training programs that focus on healthcare, hospitality and retail sectors, while the rest are enrolled in programs on computer and automobile training. Table A1 provides the demographic description of our sample. In Columns 2 and 3, we also compare our study sample to a nationally representative sample of the 68th Round of the National Sample Survey (NSS) conducted in 2011-12.[11] 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.

Details on the placement process: An important aspect of the training program, central to this paper, is that placements and allocation of job interviews to trainees is done by Placement Managers. These managers usually have a relationship with trainees. While they are not teachers in the course themselves (except in 2 cases), they are also the Center Managers and therefore responsible for enrollments, evaluations, and monitoring of trainees over the course of their training.

The placement process is typically the same across managers, with minor deviations. Managers usually start by searching for vacancies around two weeks prior to the completion of a training program. This is an important task given that most centers in our sample are located in rural areas of Uttar Pradesh and local vacancies are hard to come by. Using a list of previous employers, a manager reaches out to them to inquire about potential vacancies. These jobs are usually low/medium skilled ones (like stocking shelves, waiters, delivery boys, etc.) and employers are typically chains that have many small operations (like restaurants, coffee shops, etc.) Firms do not typically contact managers on their own. Managers first schedule interviews with an employer and then reach out to students in class who they think are a good fit for the job. Qualitative surveys indicate that managers explicitly do not want to provide students with multiple interviews, since it might confuse students in making choices and given the dearth of jobs to begin with, helps them provide other students with interviews as well. This is true in our data as well. Conditional on getting an interview, 80% of students receive only one interview, and over 90% receive at most two interviews. Similarly, students are free to reject these interviews as well, but given the excess demand for these jobs it is very unlikely that they do.[12]

3. Eliciting preferences over jobs

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

3.1. Hypothetical choices

Job aspirations

In a survey implemented during the first week of the training program, trainees were asked about their aspirations with regard to employment after the training program. We focused specifically on four aspects of a job that from other accounts, were important for trainees: employment sector, location, salary and whether there was Provident Fund (PF).[13] 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 on where they *aspired* to work after the training program. We then create a binary variable that takes the value 1 for the sector that an individual aspires to work in and 0 for others and report the results in Panel A of Table A2. 72% of trainees reported aspirations to work in the Healthcare, Banking and Retail sectors. Next, keeping in mind their qualifications, trainees were asked to describe the characteristics (salary, location and PF) of their *ideal* private sector job.[14] From Panel B of Table A2, trainees reported a desired salary of Rs. 15,036 on average,[15] with 98% reporting a preference for a job with PF. From Panel C, only 18% trainees in Uttar Pradesh aspired to get a job in the local area, while 74% aspired to get a job in a major city in Uttar Pradesh, and only 8% were willing to move outside of the state (mainly to Delhi or Mumbai, both large metropolitan cities). For the trainees in Delhi, 97% wanted a job in Delhi.

[10] http://theskillsacademy.in.

[11] 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.

[12] Qualitative interviews with students suggest that they are indeed concerned about not getting interviews at all and there is sufficient uncertainty in the process that they rarely reject interviews.

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

[14] The exact question was "Given your qualifications and background, what is the ideal job that you would want". It was constructed intentionally to nudge respondents to think about what salary they would want among those they can realistically expect to get. These questions are purely descriptive and not used for the subsequent analysis.

[15] There is variation in the expected salary across states with an average of Rs. 24,373 in Delhi and Rs. 12,978 in U.P. When we compare this to the salary actually got after placement, the average salary is Rs. 8176 in Delhi and Rs. 6622 in U.P. This difference is statistically significant at the 0.01 level.

Job priorities

In the same survey as above, trainees were asked directly about their preferences over six different job characteristics,[16] by asking them to distribute a hundred points across them. Column 2 of Table A3 reports the average points allocated by trainees to a job characteristic, Columns 4 and 5 report the values for males and females respectively and Column 6 reports the p-value that tests the statistical difference them. Salary, location and job title/designation were the three most important characteristics for trainees in a job. They were 1.5 to 2 times more important in magnitude than other job characteristics like job security, social status and nature of work. The only significant difference across men and women is with respect to location, which not surprisingly in the Indian context, is more important for women than for men.[17]

3.2. Incentivized elicitation of preferences from real jobs

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

To begin, we first generated a list of sector-specific jobs by varying the job characteristics that trainees reported as important in the hypothetical activity above: salary, location, designation and social security. The idea of this exercise was to vary job characteristics to generate jobs that closely resembled those available to trainees after the training program. Salary was varied between low, medium and high categories (based on terciles). 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.[18] Taking all possible combinations across the four characteristics would produce 36 jobs. However, we wanted to ensure that the jobs presented closely resembled jobs that the trainees could potentially get after their training program. So after looking at the history of previous jobs offered in each sector, the list of 36 jobs was narrowed down to the 11 most realistic jobs (see Figures A2 and A3).[19] To further enhance the authenticity of this exercise, it was timed to coincide with the actual placement period of the training program (usually in the last week of training).

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

Specifically, within every training batch, half of the trainees (chosen at random) were told that the job ranking activity was for research purposes, and there was a very low likelihood that the job ranking exercise would influence the interviews they would get. The other half were told that there was a very high likelihood that their job rankings would determine the interviews they would get. In both cases, because of our partnership with Skills Academy, the description was factually correct. One challenge we faced in implementing this exercise however, was that since it was conducted in the last week of the training program (just prior to placements), there was irregular attendance in the training program. Therefore, despite multiple visits to the training center, we were only able to conduct the exercise for 338 trainees (63% of the sample).[20]

Turning to the results, we find a substantial heterogeneity in trainee preferences over the *same* set of jobs (Columns 2–4 of Table 1). For each of the 11 jobs, we calculate the fraction of trainees who placed a job in the Bottom Three (Column 2), in the Middle i.e. between 4–8 (Column 3) or in the Top Three (Column 4). For example, around a third of the trainees put Jobs 2, 3, 4, 8 and 9 in their bottom three jobs, but around 20% put them in the top three. The reverse is true for Jobs 6, 10 and 11. In other words, not everyone wants the same jobs. This is why there are potentially large welfare gains from reallocating jobs based on preferences. Finally, we see no difference in the rank given to a job based on if a trainee was allocated to the Low or High likelihood group (Columns 5–7 of Table 1). The differences are both small in magnitude and nowhere near statistical significance. Going forward, we therefore assume that these rankings reflect the true underlying preferences that trainees have over these jobs.[21]

Lastly, we examine whether the job priorities that trainees report in Section 3.1 are consistent with the revealed preferences from this job ranking exercise. This would increase our confidence in using trainee preferences in interpreting the experimental results and placement outcomes discussed subsequently. We provide a brief discussion here, with a more detailed discussion in Appendix Section D. First, we use trainees' points allocated across job characteristics (in Section 3.1) to weight the characteristics in the 11 jobs that trainees rank. This allows us to generate a "hypothetical" ranking of these jobs using trainees' job priorities, which we can then compare to the job ranking exercise described above. As reported in Appendix Figure D1, we find a strong, positive correlation between these two ranking distributions for the same jobs. We also use an alternate approach that imputes the relative weights that trainees would have allocated across job characteristics to rationalize their job rankings. Specifically, using the job rankings discussed above, we estimate a rank-ordered logit specification at the trainee-level to calculate the relative weight that an individual gives to a job characteristic k (relative to salary). We then compare it to the relative points allocated by the trainee to a

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

[17] Based on a more incentivized elicitation of preferences described in Section 3.2, we calculate location compensating differentials for men and women in Section C in the Appendix. Our results suggest that keeping all other job characteristics the same, men would need a 1% and 55% increase in their real monthly salary to compensate for their disutility for accepting a job in another city in Uttar Pradesh (their state of residence) or other metropolitan cities in the rest of India respectively. This differential is 20 and 136% respectively for women.

[18] The variation in job characteristics is summarized in Table A4. 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.

[19] In general, there are many things that job-seekers could care about in a job. But given the type of employers in our sample (see Section 2.2), it was unlikely that the managers or the trainees knew much about the actual team in each location, or the job amenities when allocating interviews. The information they could realistically have had is likely to be about this rather limited set of dimensions that we use. Moreover, from Table A3, these dimensions are indeed the ones that the trainees do care most about in a job as well. In what we will show later, the managers seem to lack information even on these limited dimensions, which has consequences while matching workers to job.

[20] Table A5 shows no systematic difference in the observable characteristics of trainees who were absent on the days that this activity was conducted. For the sample of trainees for whom we do have the rankings, trainees assigned to Low and High salience groups were similar in observable characteristics (Table A6).

[21] In a related exercise, we also find no difference in the average probability or the likelihood distribution that trainees from either group (High and Low salience) put a Job X (1 to 11) in the Bottom, Middle or Top category.

Table 1
Job ranking and strategic reporting.

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

Notes: For each of the 11 jobs, Columns (2)–(4) report the fraction of trainees who ranked a job amongst the Bottom three, rank 4–8 and Top three jobs. Columns (5) and (6) report the average rank that is given to a job by trainees in the Low and High Salience groups. A higher rank indicates more preference. Column (7) reports the p-value for a t-test that tests the statistical difference between Columns (5) and (6).

Table 2
Manager knowledge of trainee preferences.

	Measure of knowledge	Reported rank	Random process	Perfect knowledge
	(1)	(2)	(3)	(4)
1.	Rank of Manager's Top Choice	7.2^{***}_{***}	5.5	11
2.	Average Rank by Trainee	6.76^{***}_{***}	6	10
3.	Most Preferred by Trainee	9.38^{***}_{***}	8.25	11
4.	Correlation b/w Preferences	0.1^{**}_{***}	0	1

Notes: Each row in Column (1) is a different measure of the manager's knowledge of trainee preferences. 'Rank of Manager's Top Choice' is the actual trainee preference for the job that the manager thinks the trainee will like the most. 'Average Rank by Trainee' is the average trainee preference across the three jobs chosen by the manager. 'Most Preferred Job by Trainee' is the rank of the most preferred job by the trainee among the three jobs chosen by the manager for her. 'Correlation b/w Preferences' is the correlation between the preference ordering of trainee and the manager. Column (2) reports the average value for each variable of interest. Columns (3) and (4) calculate the average value for a measure if the manager had with no or with perfect information of trainee preferences respectively. The stars on the top and bottom row for each value are the results from a t-test that compares the value in Column (2) to those in Column (3) and (4) respectively. * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$ respectively.

job characteristic k (relative to salary) from Section 3.1. Once again, we find that the job rankings are strongly correlated with the job priorities reported by trainees, increasing our confidence in the measurement of these rankings reflecting true underlying job preferences of trainees.

4. Do managers know what they need to know?

In this section, we identify the particular matching friction that we emphasize in the paper: the fact that placement managers, who are primarily responsible for matching trainees to jobs, do not necessarily know the preferences of the people that they are placing, thus resulting in inefficient matches. We first examine if managers were aware of trainee preferences over jobs. To do this, we used the *same* list of 11 jobs that was provided to trainees (in Section 3.2), and for *each* trainee, we asked managers to list (in order of preference) three jobs out of the 11 jobs that the trainee would like to work in. We then use this information in two ways: first, in Section 4.1, we construct multiple measures of "how well" a manager knows her trainee's preferences, and show that while they do better than just random guessing, it is far from perfect. Second, in Section 4.2, we examine whether managers do systematically better at predicting the preferences of some job-seekers as opposed to others based on characteristics of the job-seekers, jobs, managers, training programs, centers, etc. Lastly, in Section 4.3, we explore whether the information asymmetry is particularly salient along certain job characteristics as compared to others.

4.1. How well do managers know trainee preferences?

Using the manager and trainee preferences, we construct four measures of "how well" a manager knows her trainee's preferences. As a benchmark, we can compare each of our measures (described below) to two hypothetical scenarios: one where the manager responds with a random list of jobs, and one where the manager has perfect knowledge of trainee preferences and responds based on that. The results for this activity are reported in Fig. 2 and Table 2. We now discuss the four measures in detail below:

1. *Measure #1:* We consider a job that was picked by the manager as the best job for a trainee and report the rank provided by the trainee for that same job. If it were done randomly, the average rank should be close to 5.5 and if the manager knew the preferences of the trainee perfectly, this should be 11. We find that the average is 7.2 using actual trainee preferences (Row 1, Table 2). This does significantly better than a random process but significantly worse than the case if preferences were known perfectly.
2. *Measure #2:* We take all the three jobs chosen by the 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. Random choice would generate an average rank of approximately 6 while in the perfect information case it should be 10. The average observed in the data is 6.76 (Row 2, Table 2), which again does better than a random allocation, but worse than perfect knowledge.
3. *Measure #3:* Among the three jobs chosen by the manager, we take the rank of the most-preferred job by the trainee. Random choice would give us an average rank of 8.25 across trainees, and if preferences were known perfectly, this should again be 11. The average observed in the data is 9.38 (Row 3, Table 2), which is statistically better than a random process and worse than perfect information.
4. *Measure #4:* We consider the correlation between the rank orderings of the manager and the rank ordering of the trainee. With random choice, this correlation should be 0, while with perfect information, it should be 1. The average in the data is 0.1 (Row 4, Table 2), which is again better than a random process, but far worse than perfect information.

The above analysis suggests that irrespective of which measure we use, managers do better than choosing jobs completely at random, but is nowhere near perfect information. Furthermore, there is a considerable amount of variation in the manager's knowledge on trainee preferences as well. For example, as reported in Fig. 2, a trainee's most preferred

Table 3
Impact on interviews and job characteristics.

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

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

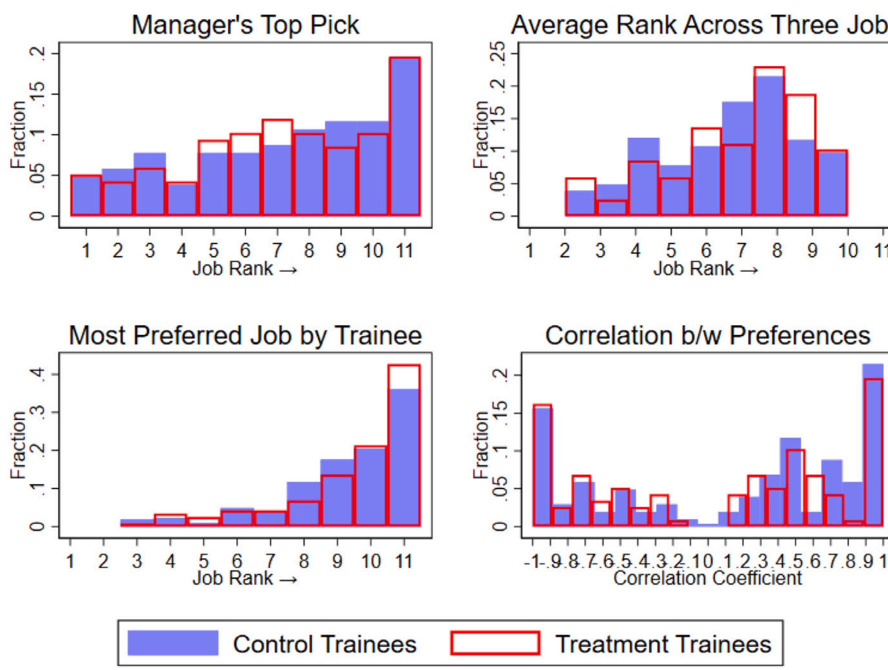

Fig. 2. Manager's knowledge of trainee–job rankings. *Notes:* The above plots histograms for the four measures of "how well" a manager knows the preferences of a trainee (discussed in Section 4.1). 'Manager's Top Pick' is the actual trainee preference for the job that the manager thinks the trainee will like the most. 'Average Rank Across Three Jobs' is the average trainee preference across the three jobs chosen by the manager for the trainee. 'Most Preferred Job by Trainee' is the rank of the most preferred job by the trainee among the three jobs chosen by the manager for her. 'Correlation b/w Preferences' is the correlation between the preference ordering of trainee and the manager for the three jobs chosen by the manager. The Job Rank goes from 1 to 11 where 11 is the most-preferred job by the trainee. We report the distribution separately for trainees in the Control (blue) and Treatment (red) groups. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

job is in one of the three jobs picked by the manager in only 40% of the cases. Furthermore, in 20 percent of cases, what the manager thinks is the best job for the trainee, coincides with the trainees most preferred job as well, but in another 20% of cases, it is actually among her bottom three jobs.

4.2. Whose preferences does the manager not know?

Are managers systematically better at predicting the preferences of some trainees more than others? In particular, are there trainee, job, trade or center characteristics that could help the manager predict trainee preferences? We examine this question in two ways. First, we regress each of the four measures M_i from the previous section for a trainee i on a set of trainee characteristics X_i to estimate the following specification:

$$M_i = \alpha_b + \beta X_i + \varepsilon_i \quad (1)$$

where α_b are batch fixed effects that take into account all observable and unobservable characteristics about a certain center, area, trade or manager that would affect these preferences. We consider four characteristics of an individual, namely: (a) Female, which is a binary variable that takes the value 1 if the individual is female and 0 otherwise; (b) Age; (c) Years of education of the individual; (d) Backward Caste, which is a binary variable that takes the value 1 if the individual is from a SC, ST or OBC caste and 0 otherwise. As reported in Table A9, we do not find any evidence that managers are potentially better at predicting the preferences of their trainees based on these characteristics.

A concern with using this specific vector of trainee characteristics is that is may not be predictive of preferences, or there might be other second- and third-order interactions (lower caste males, young women, etc.) or job, center, batch characteristics that would be observable to managers and predictive of trainee preferences. We therefore adopt a more general approach here. Specifically, we take a rich set of candidate characteristics, namely: female, education, age, caste category, religion and the year of education of the mother and father. We then generate double and triple interaction terms of female and caste category with individual education and age respectively. Put together, this gives a set of 40 trainee characteristics. We then estimate an unpenalized adaptive LASSO regression to select those trainee characteristics that are most predictive of her job preferences. As reported in Table A10, the procedure selects 13 variables. We then estimate a linear regression of job preferences (measured by the rank given by a trainee i to a job j) on the vector of these 13 variables. The R^2 (Column 3, Table A10) of this regression is just 1.2%, indicating that these characteristics are able to explain very little variation in trainee preferences.

We now redo the analysis by adding a vector of rich job characteristics (a dummy variable for each job) interacted fully with female and caste-categories to the previous vector of trainee characteristics. Note that the dummy variables for jobs capture all observable and unobservable characteristics of a job. Furthermore, the interaction terms allow them to vary by gender, caste or both. From Row 2 of Table A10, the LASSO estimation selects 54 variables out of the possible 183 (Column 3), which explain 13.1% of the variation in trainee preferences (Column 4). Lastly, we further add dummy variables for each center-batch to the vector of trainee and job characteristics from above. These dummy variables therefore capture all observable and unobservable characteristics of a trade, sector, center or manager that could help predict trainee preferences. From Row 3 of Table A10, the LASSO estimation selects 54 variables out of the possible 206 (Column 3), which explain 13.2% of the variation in trainee preferences (Column 4).

Put together, the above analysis indicates that a rich set of trainee, job, center, trade and manager characteristics–proxying for the information set of a manager–cannot explain a majority of the variation in trainee preferences.

4.3. What do managers not know?

Having established the information asymmetry in the knowledge of trainee preferences by managers, we now examine whether this lack of information is particularly salient along specific job dimensions. In particular, we take the three most-preferred jobs for a trainee as well as the three jobs that the manager thinks a trainee would like. For each job characteristic (location, salary, etc.) we then ask what fraction of top three jobs reported by a trainee had these characteristics, and then compare it to the counterpart for the manager. For example, in Figure A4a, we see that only 18% of trainees have their top three jobs in the 'Same District' (gray bars). This is in sharp contrast to the managers (blue bars), who think that 49% of trainees would want their jobs in the 'Same District'. Consequently, we see that there is a mismatch in knowledge of preferences for jobs in the 'Same State' (49% for trainees and 37% for managers) and 'Outside the State' (33% for trainees and 14% for managers). Since local jobs are also lower paying than jobs in bigger cities, we see the same pattern on salary preferences (Figure A4b). On the other hand, trainee preferences and their knowledge by managers seem better aligned on job activity and PF (Figures A4c and A4d).[22]

To summarize, the above analysis identifies the friction that is at the heart of this paper: placement officers, who are directly and completely responsible for the matching job-seekers to jobs, do not seem to know the preferences of many of their job-seekers.

[22] A qualitative survey with managers revealed that job-activity is usually correlated with the trade of the training program. So, training programs in the retail sector for example, are very likely to place students in 'active' jobs of stocking shelves or as delivery boys. On the other hand, training programs for call centers would most likely offer 'phone' jobs. Therefore self-selection into different types of training makes it easier for the managers to know trainee preferences along these dimensions.

5. The impact of informing managers

After eliciting preferences of trainees across jobs and establishing the manager's lack of knowledge of these preferences, we now describe the randomized control trial associated with informing managers about trainee–job preferences and the consequences it had. Before we proceed, note that the fact that the lack of managers knowledge on trainee preferences is not by itself evidence that they will (or even should) use the information we provide them since they may have other information, say about employers' preferences, which may also be relevant. However, as we will subsequently show in this section, they do make use of the information on preferences that we provide them with. This still leaves open the possibility that managers are in fact wrong to use our information and should have stuck to what they were doing, since they have information that we do not have. We address this in Section 7, where we find no evidence that the matching is worse for the treated trainees. Their labor market outcomes are, if anything, better.

5.1. Intervention details

As discussed in Section 2.2, managers usually contact various firms for job vacancies towards the end of the training program, and then decide how to allocate them across trainees. Our intervention aimed at reducing the asymmetry o information on job preferences between the trainees and managers as follows: trainees in each batch were randomly allocated to one of two groups. For the first group, henceforth the Treatment group, we provided a description of the job characteristics for the top four jobs ranked by the trainee to the manager (see Figure A5 for two examples). For the second group, henceforth the Control group, no trainee preferences were shared with the manager. As reported in Panel A of Table A7, the trainees in the two groups were similar on observable characteristics. Moreover, as reported in Panel B as well as Fig. 2, managers also had similar knowledge on job preferences of trainees in the two groups, as captured by the four measures discussed in Section 4.1.

5.2. Impact on the number of interviews

We begin by examining whether the treatment had any effect on trainees getting more interviews or a different set of interviews by estimating the following regression:

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

where T_i is a binary variable that takes the value 1 if the trainee was in the Treatment group and 0 otherwise. α_b are batch or strata fixed effects. X_i are a set of trainee characteristics like age, gender, education and dummy variables for if the trainee is currently a student or of a lower caste.[23] We report the results in Columns 1–3 of Table 3. We consider the following outcome variables: the number of interviews received by trainee a i (Column 1), the number of interviews conditional on getting at least one (Column 2), dummy variables that equal 1 if the trainee gets an interview for (a) any job and (b) at least one job in her four most-preferred jobs (Columns 3 and 4 respectively).[24] As reported in Columns 1–3, we find no differential impact of the treatment on the number of interviews received.

[23] The results are robust to controlling for the trainee's average job rank for the three jobs chosen by the manager (Measure #2 in Section 4.1) to account for the fact that a manager may have more information for certain trainees.

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

5.3. Quality of interviews: Data challenges

Given that there is no effect on the number of interviews, it is somewhat easier to interpret the next set of results, which are about the quality of the match. We examine whether treated trainees were matched to interviews that they preferred more. There were two challenges that we encountered with the placement data: first, in the set of 11 jobs that were ranked by the trainees, we had varied the designation of the job (between active and desk jobs). However, most of the firms that candidates were actually matched to did not specify the type of job that they would place the trainee in, and so we could not match this dimension of preferences with the data. We therefore take the 11 jobs and average the rank over the designation dimension. This leaves us with 8 jobs for every trainee that now only vary in terms of salary, location and Provident Fund.[25]

The bigger challenge was that if we took the complete set of combinations along the three dimensions (salary, location and PF) we would have 18 potential jobs. However, as discussed earlier, to make the activity more realistic, we dropped some jobs based on the previous placement experience of Skills Academy. In the placement data however, we do encounter interviews where the set of job characteristics do not correspond to the jobs ranked by trainees. Out of a total of 217 interviews that we have in our data, we are able to perfectly match around two-thirds of the interviews (141 to be exact) with those in the job ranking list. For the remaining interviews, we do not have a match (and hence we do not know the preference of the trainee). Going forward, we only consider interviews where we know the trainee preferences.[26] The last row of Table A7 shows that the number of interviews that we were able to match with preference rankings is not correlated by treatment assignment, as one would expect.

5.4. Impact on match quality

Since the intervention involved providing a manager with information on the four most preferred jobs of the trainee, we now examine the impact of this intervention on two outcome variables: (i) a dummy variable on whether a trainee received at least one interview for her four most preferred jobs and (ii) the (normalized) average preference across all interviews received by the trainee. We then re-estimate Eq. (2) and report the results in Columns 4 and 5 of Table 3. We see that treatment trainees are 10.7 pp (48.4%) more likely to get an interview for one of their four most preferred jobs. More generally, as reported in Table A11, we find that treated trainees have a higher probability of getting an interview for their top X jobs, where X varies from 1 to 5. Lastly, the average placement quality, as measured by the normalized rank across interviews, is 0.26σ higher for the treatment trainees as compared to control ones.[27]

6. Spillovers of matching

Given that treatment and control job-seekers in our experiment were competing for the same pool of interviews, we cannot directly conclude that our intervention increased aggregate welfare. In particular, the intervention may have actually made things worse on average when one includes the Control group. This is because we gave managers information on preferences of half the trainees they had to assign to interviews, while saying nothing about the others. This can easily lead a manager to move to an allocation which is worse on average, and from one that is in the core to one which is not.[28] To illustrate this with a simple example, consider three jobs: $1, 2, 3$ and three job-seekers: a, b, c. Let their preferences be: $\{(1P_a3P_a2),(1P_b2P_b3)(3P_c2P_c1)\}$. In the original allocation, the manager has some very noisy information about b's top preference and nothing else. Based on that, she chooses the allocation $\{a \rightarrow: 3; b \rightarrow 1; c \rightarrow 2\}$. b gets what the manager's best information says should be her top choice. Now suppose the manager is given very precise information about $a's$ preference and decides that she has no reason not to give a his top preference and then switches b to job 3, to generate the allocation $\{a \rightarrow 1; b \rightarrow 3; c \rightarrow 2\}$. 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 reduces by one, while those with their top preference is still one.

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

To make progress however, we take two approaches to evaluate the importance of spillovers. First, we take a simple approach: we hypothesize that if managers shrink their assessments of individual-specific preferences towards group mean ones, it will likely induce congestion in job allocation. As managers receive more accurate information on preferences of Treatment trainees, the intervention can then reduce this congestion to allow for more efficient matches without much spillovers. A key testable pattern under this hypothesis would be to find that trainee preferences as reported by managers are less dispersed than trainees' self-reported preferences. We test this "shrinkage hypothesis" in Section 6.1.[29]

Our second approach is more theoretical in nature. Note that our main challenge is that we need to predict the allocation of interviews between the Treatment and Control trainees in the absence of our intervention. For this, we would first need to reliably model the manager's decision rule based on the observed allocation of interviews (Section 6.2). Next, assuming that this rule is a reasonable approximation of how the manager actually decides, we can then generate a counterfactual allocation of interviews for individuals in Treatment and Control in the absence of the intervention. This would allow us to examine the impact of our intervention after accounting for any spillovers (Section 6.3).

6.1. Testing the shrinkage hypothesis

Let r_{ij}^m be the rank given by a manager for a job j and trainee i, and r_{ij} be the corresponding rank given by the trainee for the same

[25] It is important to note that these 11 jobs do not vary uniformly across all dimensions. Therefore, by collapsing along one dimension (job activity for instance) would not necessarily imply reducing the number of jobs by half. In fact, among the 11 jobs, there were 6 jobs that differed on the job activity margin, keeping other job characteristics equal. Therefore, once we collapse them down to 3 jobs, along with the 5 other jobs gives us our list of 8 jobs.

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

[27] As discussed in Section 4.1, there is a lot of variation in the knowledge managers have about trainee preferences. However, we do not find any heterogeneous treatment effects along this dimension.

[28] A core allocation is where two trainees cannot swap interviews with each other to make both better off.

[29] We would like to thank an anonymous referee for this suggestion.

Fig. 3. Testing the shrinkage hypothesis. *Notes:* The above figure plots the CDF of the ratio of the dispersion in managers' assessment of trainee preferences (σ^m) as compared to self-reported trainee preferences (σ). The gray line plots the distribution for each batch separately, while the solid black line plots the CDF after pooling across all jobs and batches.

job (where a higher rank indicates a higher preference).[30] For each job j in a batch b, we then calculate the dispersion in the managers' and trainees' ranks as measured by the standard deviation i.e., we calculate $\sigma_j^m = S.D.(r_{ij}^m)$ and $\sigma_j = S.D.(r_{ij})$. Shrinkage would imply $\sigma_j^m < \sigma_j$ or $\sigma_j^m/\sigma_j < 1$. Fig. 3 plots the CDF of this ratio σ_j^m/σ_j, for each batch separately in the gray lines and pooling across all batches in the solid black line. We see that across all batches, this ratio is less than 1 for over 80% of jobs. Even within a batch, this ratio is less than 1 for 60%–100% of jobs. The corresponding p-value of a Kolmogorov–Smirnov test that tests for the equality of the σ_j and σ_j^m distributions is less than 0.001 — both within a batch, as well as at the aggregate level. This indicates that the our intervention could reduce the shrinkage of the perceived trainee preferences by managers, resulting in a more efficient allocation of jobs without much spillovers.

6.2. The manager's interview allocation rule

While the evidence in the previous sub-section suggests that spillovers be less of a concern than we might have imagined, we now take a more rigorous theoretical approach to measuring potential spillovers. Specifically, we model the managers' decision rules and use them to simulate allocations in the absence of our intervention. There are three components to understanding how a manager would have allocated interviews in the absence of our intervention. First, we need to define the manager's information set i.e., her knowledge of trainee preferences both with and without the intervention. Second, we need to devise an algorithm to allocate the set of interviews across trainees (conditional on the manager's information set) and lastly, we need to examine how the simulated allocation with the intervention compares to the actual allocation that we can empirically observe. We discuss each step below.

Information set of the manager

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

Allocation mechanism for interviews

To assign a decision-rule to the manager in allocating these interviews, we assume that under each of the hypothesized information sets, she chooses allocations that are in the core i.e., allocations where two trainees cannot swap interviews with each other to make them both better off. We can then compare the predicted allocations under each information set of a manager with the actual empirical allocation to choose an information set that is most likely used by her. Our algorithm to identify these allocations is as follows: trainees in a batch are arranged in a random order and their manager sequentially allocates an interview to them from the set of available interviews. For example, after the manager allocates the first trainee her interview, the next trainee is allocated one from the remaining interviews, and so on. Note that in doing so, we assume that managers have no preferences over which trainee should get which interview.[31] Secondly, for almost all batches, there are more trainees than interviews–so any allocation rule would have multiple allocations in the core. To take this into account, we run the algorithm 25,000 times, each time ordering trainees randomly within each batch to simulate the set of allocations. We can therefore calculate the probability that a trainee i is matched to an interview for job j (denoted by p_{ij}).[32] Third, we empirically observe a few individuals in every batch getting multiple interviews. In fact, in 6 out of 21 batches, more than 15 percent of trainees get multiple interviews (see Figure A6). So unless we make further assumptions on how individuals value 'bundles' of interviews, we cannot perfectly compare the theoretical and empirical allocations (since in the simulated allocations, every individual gets only one interview). For our main results therefore, we drop these six batches, though we also show that our results are robust to including all batches as well (see Figure A7).

Results

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

[30] Note that managers only provide us a ranking for 3 jobs, while the trainees rank 11 jobs. To make the exercise comparable across trainees and managers, we set the trainees' self-reported rank for all jobs except the top three to 0. Similarly, we set the manager's self-reported rank for all jobs except the one they choose for the trainee to be 0 as well. In this way, the rank for each job j will always range between 0–3 for both trainees and their managers.

[31] It is plausible that the manager acts in the employers interest and chooses certain trainees because they fit the employers needs better. We rule that out by assumption since it most likely is not the dominant practice in our setting. Employers do not exhibit strong preferences in hiring specific candidates given the nature of the jobs — for example, around 75% of trainees who get an interview are also offered a job.

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

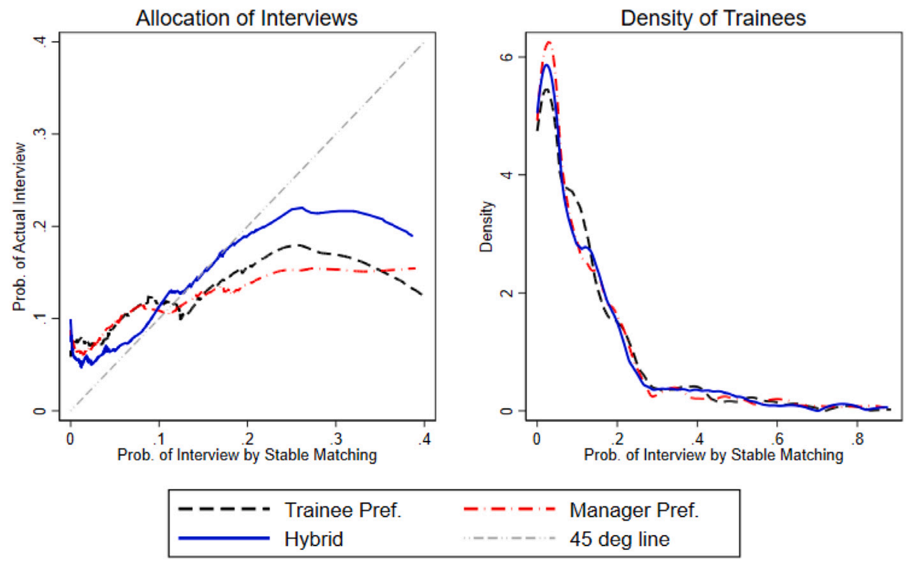

Fig. 4. Simulated and empirical allocations. *Notes:* The horizontal axis in both graphs show the probability that a trainee i gets an interview for job j as allocated by the algorithm described in Section 6 of the paper. The first graph on the left compares the simulated allocation to the empirical allocation under the three information sets of the manager: (i) Full Information, where we assume that the manager perfectly knows trainee preferences (dash black line); (ii) No Information case where we assume that the manager allocates according to what she thinks are trainee preferences (dotted red line); (iii) Hybrid Information case where we use (i) for the Treatment trainees and (ii) for Control (solid blue line). The dash-dotted gray line is the 45 degree line. The second graph on the right shows the density of trainees across the simulated probability distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4 plots this relationship. The empirical allocations should coincide with the 45 degree line if managers are allocating efficiently conditional on their information set. From the first graph, allocations under the hybrid information set do a better job at explaining the empirical allocations as compared to the other two cases. From the second graph, most trainees have relatively low values of p_{ij}, which is not surprising given the scarcity of jobs. However, it is precisely for those high p_{ij} jobs, where more information to the manager seems to be crucial in improving the allocation of interviews. This is intuitive, since it is for these jobs that being able to identify the small number of people who really want them creates a potential for a large welfare gain.

6.3. Spillovers on the control trainees

Once we have a model of the manager's decision rule, we can use it to create a counterfactual allocation of interviews in the absence of the intervention. We can use this to then examine the impact of our intervention, *after* accounting for the reallocation of interviews between the Control and Treatment trainees. To begin, under the two "no information" and "hybrid information" sets, we simulate the allocation of interviews (using the algorithm and protocol described previously). We can therefore calculate the probability (under each scenario) that a Treatment and Control trainee receives a job of rank r (where a rank of 1 is least preferred and 8 is most preferred job). We report these distributions in Fig. 5. As can be seen from the graph, there is a clear increase in the probability that a Treatment trainee gets interviews for a job that she prefers more. We then formalize this intuition by estimating the following regression specification:

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

Since our intervention provides the manager with information on the four most preferred jobs by a trainee, p_{ijm} is then the probability that a trainee i gets allocated an interview for at least one of her j most preferred jobs under an allocation rule $m \in$ {No Info, Hybrid}. T_i is an indicator variable for if the trainee is in the Treatment group, and X_i are the set of individual controls used in previous regressions. The results are reported in Table 4, where each column reports the results for the j most-preferred jobs, where j ranges from 1 (best job) to 4 (top four jobs). We find no statistical difference in the probability that a Control or Treatment trainee gets allocated an interview for their most preferred jobs under the "no-information" set. On the other hand, under the hybrid information set, Treatment trainees are 8–12 pp more likely to be allocated an interview jobs that they prefer more, while Control trainees remain unaffected (both the magnitude is small and the coefficients are statistically insignificant at conventional levels).

Table 4
Probability of getting a top-four job.

	Best job (1)	Top 2 jobs (2)	Top 3 jobs (3)	Top 4 jobs (4)
No Info. × Treat	0.0188	0.0120	0.0325	0.0217
	(0.0188)	(0.0206)	(0.0205)	(0.0191)
Hybrid × Control	−0.0131	−0.0149	−0.0179	−0.0185
	(0.0185)	(0.0218)	(0.0229)	(0.0219)
Hybrid × Treat	0.0778***	0.116***	0.108***	0.0920***
	(0.0272)	(0.0288)	(0.0280)	(0.0260)
Control, No Info. mean	0.0630	0.145	0.192	0.255
N	586	586	586	586
R^2	0.365	0.573	0.625	0.701
Ind. controls	Yes	Yes	Yes	Yes
Batch FE	Yes	Yes	Yes	Yes

Notes: The dependent variable in Columns 1–4 is the probability that a trainee i gets an interview for at least one top X jobs where X varies from 1 to 4. No-info is a dummy variable that takes the value 1 if the allocation was simulated under the no-information case, while Hybrid is a dummy variable that takes the value 1 if the allocation was simulated under hybrid information case. All regressions include individual controls and batch fixed effects. Individual controls used are the number of interviews, age, gender, years of education and manager knowledge of trainee preferences (measure #2), indicator variables for whether the trainee is a student or not and whether from a SC/ST/OBC caste category. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$ level of significance.

7. Impact on placements and employment outcomes

The above analysis is suggestive that the intervention did have an impact on improving the efficiency of the matching process. However, are trainees matched with more-preferred job interviews also more likely to accept these jobs and retain them for longer? This is after all

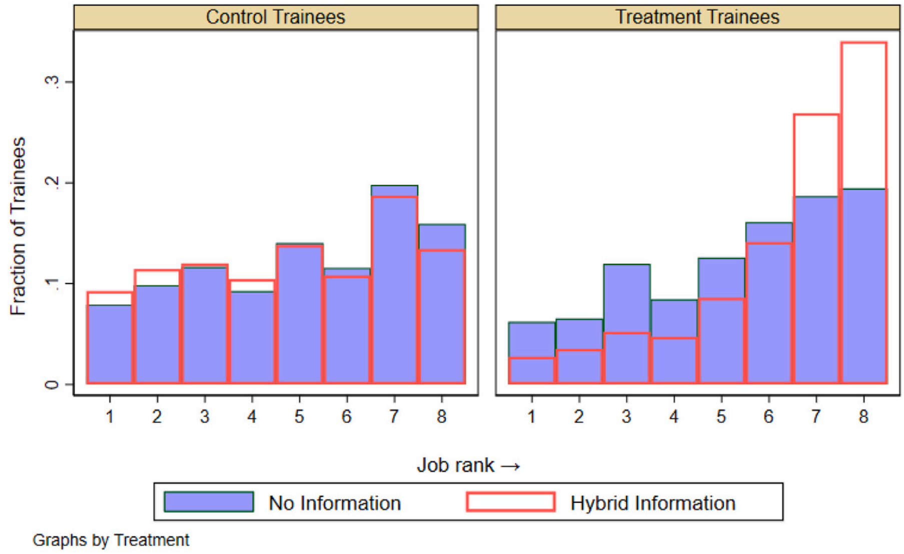

Fig. 5. Probability of Getting a Job with Rank r. *Notes:* The graph shows the probability that a trainee gets an interview for a job rank with r, where 1 is the least preferred job and 8 is the most preferred job. The histogram then shows the fraction of trainees (Control group on the first graph on the left, Treatment group on the second graph on the right) who get a job of rank r. The blue bars show the distribution under the No Information case i.e., where the manager uses what she thinks are trainee preferences to allocate interviews, while the red bars show the distribution under the Hybrid Information case i.e., where the manager uses her information set for the Control group, but (due to our intervention) has perfect information on trainee preferences for the Treatment group. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the outcome a policymaker cares about. We now examine the effect of our treatment on placement and employment outcomes in this section, which is possible since we observe: (a) the reported preference of a trainee for a job; (b) various placement and employment outcomes for every trainee–job pair.

7.1. Measuring outcomes

For a trainee i and job j, we consider three outcomes related to interviews and offers– (i) at least one interview; (ii) at least one offer; (iii) whether an offer was accepted; and four outcomes related to job retention and employment– (i) whether the trainee was employed in the same job three and six months later and (ii) whether the trainee was employed in any job three and six months later.[33] Let us denote these outcomes by y_{ij}. We then use two ways to aggregate these numbers: first is an *Unweighted* Index, where for each individual, we aggregate outcomes across all jobs to create an individual-specific placement and employment index (y_i), which takes the value 1 if $\sum_j y_{ij} > 0$ and 0 otherwise.[34] Second we create a *Preference-Weighted* Index, $y_i^{pw} = \sum_j r_{ij} y_{ij}$, where for each individual, we aggregate outcomes across all jobs *after weighting* them with the individual's ranking for that job (r_{ij}).

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

7.2. Impact of the intervention

We now turn to discussing the effects of our intervention on employment and placement outcomes in Table 5. Column 2 reports the mean of the various outcome variables for the Control trainees. We start with the *unweighted* placement and employment outcomes (Columns 3–4). Column 3 reports the raw difference between an outcome variable for Treatment and Control trainees, while Column 4 reports the regression coefficient after controlling for individual controls and batch fixed effects ($\hat{\beta}$ in Eq. (2)). We report the corresponding p-values for each statistic in parentheses below.

Turning to the results (Panel A of Table 5), 50% of trainees in the Control group received at least one interview, 36% received at least one offer and only 18% accepted an offer. However, conditional on receiving at least one interview, 72% received at least one offer, and conditional on receiving an offer, 50% of trainees accepted it. The difference between the Treatment and Control groups (Columns 3–4), is both negligible in magnitude and statistically insignificant at conventional levels.

From Panel B, 9% and 12% of Control and Treatment trainees respectively were employed in the same job three months later. However, note that conditional on accepting a job after the training program (see Panel C), 48% and 67% of trainees were likely to stay employed in the same job after three months. The difference between the two groups is large in proportional terms– 33% in Panel B and 39% in Panel C respectively, but too imprecise to be statistically significant at conventional levels. Lastly, 25% and 29% of Control and Treatment trainees were employed in any job after three months (see Panel B). Restricting our sample to trainees who accepted a job after employment (see Panel C), 72% and 78% of Control and Treatment trainees were employed in a job after three months. The difference between the two groups (16% in Panel B and 8.3% in Panel C) are again quite substantial in magnitude, but not statistically significant at conventional levels.

[33] The lower sample size is because we were able to survey 91% of our trainees after six months. As reported in Table A8, attrition from the sample is not correlated with a trainee's treatment status, or a wide range of their socio-economic characteristics such as education, work experience, caste, and parental age and education. Attrition is higher for women and younger individuals. We control for both gender and age in our analysis throughout the paper.

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

Table 5
Impact on job choice and employment outcomes.

	N	Control mean	Unweighted Diff.		Quality weighted Diff.	
			b/w T-C		b/w T-C	
			No FE	Batch FE	No FE	Batch FE
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Placement Outcomes for All Trainees						
At least one interview	293	0.5	0.02	0.02	0.27**	0.27**
			[0.78]	[0.72]	[0.02]	[0.02]
Offer received	293	0.36	0.00	0.00	0.19*	0.18*
			[0.93]	[0.97]	[0.09]	[0.10]
Offer accepted	293	0.18	0.00	0.00	0.24**	0.25***
			[0.96]	[0.99]	[0.05]	[0.04]
Panel B: Employment Outcomes for All Trainees						
Same job (3 mts)	293	0.09	0.03	0.03	0.25**	0.26***
			[0.37]	[0.40]	[0.04]	[0.04]
Employed (3 mts)	293	0.25	0.04	0.03		
			[0.47]	[0.52]		
Employed (6 mts)	266	0.19	0.08	0.07		
			[0.14]	[0.18]		
Panel C: Employment Outcomes Conditional on Accepting a Job After Training						
Same job (3 mts)	52	0.48	0.19	0.19	0.59**	0.21
			[0.18]	[0.25]	[0.04]	[0.51]
Employed (3 mts)	52	0.72	0.06	0.04		
			[0.64]	[0.75]		
Employed (6 mts)	41	0.38	0.27*	0.04		
			[0.09]	[0.84]		

Notes: Panels A and B report the placement and employment outcomes for all trainees respectively. Panel C restricts the sample to trainees who accepted a job after completing their training program. Column 2 reports the mean for the Control group. Columns 3–6 report differences between the Treatment and Control group. Columns 3–4 report the unweighted differences, while Columns 5–6 report the same outcomes after weighting them by trainee preferences and normalized to have mean 0 and S.D. 1 for the Control group. Columns 3 and 5 report the raw difference, while Columns 4 and 6 report differences after accounting for batch fixed effects and individual controls. p-values are reported in square parentheses. * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$ level of significance.

Turning to a slightly longer term horizon, six months after initial placement, only 3 trainees (across both groups) stayed in the same job indicating limited persistence of staying in the same job.[35] However, as reported in Panel B, Treatment trainees were 8 pp (or 42%) more likely to be employed in any job ($p = 0.14$). Among the trainees who accepted a job after training, 38% of Control trainees were employed in a job and trainees in the Treatment group were 27 pp (or 71%) more likely to be employed in any job ($p = 0.09$).[36]

While there is no difference in the probability of getting an interview or taking a job, there seems to be some effect on job retention (though its not quite significant). The preference weighted index combines this with quality differences in the interviews and the jobs. As discussed before, for each individual i, we create a preference-weighted outcome y_i^{pw} that aggregates outcomes across jobs by weighting them by an individual's preference for it. To make comparisons between the Control and Treatment groups easier to interpret, we normalize y_i^{pw} to have mean zero and standard deviation 1 for the Control group trainees. As reported in Columns 5 and 6 of Table 5, Treatment trainees had 0.27σ better quality of interviews, 0.19σ better quality offers and 0.24σ higher acceptance of these offers. All of these results are statistically significant at conventional levels and robust to controlling for batch fixed effects. Lastly, looking at the persistence of staying in a job in Panel B, treated trainees were 0.25σ more likely to stay in this better quality job three months after placement ($p = 0.04$), though the effect does not persist six months after placement because (as discussed earlier) almost no one stayed in the same job for six months.

Our job ranking exercise in fact makes a broader point: if we are able to meaningfully capture trainees' underlying preferences, getting better (more-preferred) interviews should – irrespective of a trainee's treatment status – translate into better placement and employment outcomes more generally. To examine this, we estimate the following regression:

$$y_i^{pw} = \alpha_b + \beta \text{Interviews}_i^{pw} + \gamma X_i + \varepsilon_i \tag{4}$$

where: Interviews_i^{pw} and y_i^{pw} are the preference-weighted interviews, and placement and employment outcomes for trainee i respectively. We include individual controls (X_i) and batch fixed effects in line with the analysis throughout the paper. The results are reported in Appendix Table A12. From Column (3), we see that trainees who received better quality interviews were more likely to receive better offers, accept them, and retain them in the short run (3 months), but not in the longer run (6 months). Given that (as discussed earlier) our intervention does not impact the number of interviews, but the type of interviews trainees get, treatment can potentially be used as an instrument for the probability of getting "better" interviews. As we already know from Table 5 and Column (2) of Table A12, our intervention significantly increased the probability of getting more-preferred interviews. The resulting 2SLS results (Column 4) are qualitatively similar to the OLS results (Column 3) — trainees with more preferred interviews did causally have better employment and (short-run) placement outcomes.

How large is the magnitude of reducing this initial matching friction? We do a quick back-of-the-envelope calculation to monetize this treatment effect of 0.27σ of receiving better quality interviews by calculating compensating differentials (see details in Appendix Section C). To elaborate, we standardize the preference ranking of a trainee i for a job j to have mean 0 and standard deviation 1 and estimate Equation (5). $1/\hat{\gamma}$ (equal to Rs. 6950) can be interpreted therefore, as

[35] These characteristics seem to be a consistent feature across other similar labor markets, such as those in Ethiopia (Abebe et al., 2023; Blattman et al., 2019) and Uganda (Alfonsi et al., 2020; Bassi and Nansamba, 2022).

[36] We are unable to distinguish the different labor demand- and supply-side reasons for the job retention and employment transitions since we do not have objective data on why trainees quit their jobs such as quits, fires, non-renewals, etc.

the (average) salary of a job that is 1σ more preferred. Our treatment effect of 0.27σ is therefore equivalent to a Rs. 1876 higher salary, which is 20.8% of the maximum monthly salary among all jobs in the job elicitation exercise (Section 3.2), or 26.8% of the average monthly salary for jobs for which trainees receive an interview.[37]

To summarize the above discussion, our results indicate that trainees in the Treatment group were more likely to get interviews for jobs they prefer, more likely to accept these jobs, and this initial matching friction is substantial in magnitude (20%–25% of the average monthly salary). More generally as well, our results indicate that matching trainees to more preferred interviews (irrespective of treatment outcomes) did improve their placement and (short-run) employment outcomes.[38]

8. Conclusion

This paper identifies an important potential source of mismatch in the Indian labor market — that intermediaries (placement managers in our context) who are responsible for matching job-seekers to jobs do not know the preferences of these job-seekers and therefore assign them to the "wrong" jobs. We provide evidence for this mismatch using the placement process for a large vocational training firm in India and examine the extent to which provision of information on preferences can lead to a better allocation of interviews, jobs and employee welfare. We see this paper as a part of a larger research agenda of understanding search costs and mismatch in the labor market and ways to reduce them. While the literature has largely emphasized externalities and incentive problems, we show an example where the benefits are internal to the firm and the firm has strong incentives to get it right, but the outcome is nevertheless inefficient in the sense that some easily gathered information could lead to a much better allocation. Perhaps managers overestimate just how difficult it is to learn something useful, in the spirit of a literature on the internal management of firms that shows how managers do not always take what appear to be quite easy steps to enhance their productivity under normal circumstances, but switch to doing so when they are suggested by outside consultants (Bloom and Van Reenen, 2007, 2010; Bloom et al., 2013).

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

Lastly, it may be important to start a culture of unpaid internships in firms for high school students so that they can learn what they like–the high quit rates that we see after placement suggest that they often do not know what they are getting into. It is important to try to persuade the youth to be more realistic about their employment options, possibly by highlighting the importance of getting started early.

CRediT authorship contribution statement

Abhijit V. Banerjee: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Gaurav Chiplunkar:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2024.103330.

References

Abebe, Girum, Caria, A Stefano, Fafchamps, Marcel, Falco, Paolo, Franklin, Simon, Quinn, Simon, 2021. Anonymity or distance? Job search and labour market exclusion in a growing African City. Rev. Econ. Stud. 88 (3), 1279–1310.
Abebe, Girum, Caria, Stefano A, Fafchamps, Marcel, Falco, Paolo, Franklin, Simon, Quinn, Simon, Shilpi, Forhad Jahan, 2023. Matching frictions and distorted beliefs: evidence from a job fair experiment. Econ. J. Conditionally Accepted.
Abel, Martin, Burger, Rulof, Piraino, Patrizio, 2020. The value of reference letters: Experimental evidence from South Africa. Am. Econ. J.: Appl. Econ. 12 (3), 40–71.
Acemoglu, Daron, 1996. A microfoundation for social increasing returns in human capital accumulation. Q. J. Econ. 111 (3), 779–804.
Acemoglu, Daron, 1997. Training and innovation in an imperfect labour market. Rev. Econ. Stud. 64 (3), 445–464.
Alfonsi, Livia, Bandiera, Oriana, Bassi, Vittorio, Burgess, Robin, Rasul, Imran, Sulaiman, Munshi, Vitali, Anna, 2020. Tackling youth unemployment: Evidence from a labor market experiment in uganda. Econometrica 88 (6), 2369–2414.
Altmann, Steffen, Falk, Armin, Jäger, Simon, Zimmermann, Florian, 2018. Learning about job search: A field experiment with job seekers in Germany. J. Public Econ. 164, 33–49.
Banerjee, Abhijit V., Sequeira, Sandra, 2023. Spatial mismatches and imperfect information in the job search. J. Dev. Econ. 164.
Bassi, Vittorio, Nansamba, Aisha, 2022. Screening and signalling non-cognitive skills: experimental evidence from Uganda. Econ. J. 132 (642), 471–511.
Beam, Emily A., 2016. Do job fairs matter? Experimental evidence on the impact of job-fair attendance. J. Dev. Econ. 120, 32–40.
Belot, Michele, Kircher, Philipp, Muller, Paul, 2019. Providing advice to jobseekers at low cost: An experimental study on online advice. Rev. Econ. Stud. 86 (4), 1411–1447.
Bertrand, Marianne, Hsieh, Chang-Tai, Tsivanidis, Nick, 2021. Contract Labor and Firm Growth in India. Technical Report, National Bureau of Economic Research.
Blattman, Christopher, Dercon, Stefan, Franklin, Simon, 2019. Impacts of Industrial and Entrepreneurial Jobs on Youth: 5-year Experimental Evidence on Factory Job Offers and Cash Grants in Ethiopia. Technical Report, National Bureau of Economic Research.
Bloom, Nicholas, Eifert, Benn, Mahajan, Aprajit, McKenzie, David, Roberts, John, 2013. Does management matter? Evidence from India. Q. J. Econ. 128 (1), 1–51.
Bloom, Nicholas, Van Reenen, John, 2007. Measuring and explaining management practices across firms and countries. Q. J. Econ. 122 (4), 1351–1408.
Bloom, Nicholas, Van Reenen, John, 2010. Why do management practices differ across firms and countries? J. Econ. Perspect. 24 (1), 203–224.

[37] While the average impact is 0.27σ, the 95% confidence interval ranges from 0.05σ to 0.49σ. This translates into a range of Rs. 347 to Rs. 3405, or 3.8%–37.7% of the maximum monthly salary, and 4.9%–48.6% of the average monthly salary.

[38] We asked trainees the reason for quitting jobs to understand whether these job transitions arose from quits, fires, non-renewals, etc. Since these were self-reported measures, no one reported being fired from a job. However, distance to job (39.2%), not enjoying the work they do (30.7%), and low salary (28%) were the most commonly reported reasons. This suggest that trainees appear to update their beliefs and preferences after being exposed to certain job environments, labor markets, and urban living conditions. While rigorously documenting how this happens is beyond the scope of this paper, it is definitely a promising next step for future research given that our findings are not unique to South Asia alone, but extend to various other countries in Africa and South-East Asia as well.

Card, David, Kluve, Jochen, Weber, Andrea, 2010. Active labour market policy evaluations: A meta-analysis. Econ. J. 120 (548).

Card, David, Kluve, Jochen, Weber, Andrea, 2018. What works? A meta analysis of recent active labor market program evaluations. J. Eur. Econom. Assoc. 16 (3), 894–931.

Chaurey, Ritam, 2015. Labor regulations and contract labor use: Evidence from Indian firms. J. Dev. Econ. 114, 224–232.

Cheung, Maria, Egebark, Johan, Forslund, Anders, Laun, Lisa, Rodin, Magnus, Vikström, Johan, 2019. Does job search assistance reduce unemployment? Experimental evidence on displacement effects and mechanisms.

Chiplunkar, Gaurav, Chaurey, Ritam, Soundararajan, Vidhya, 2023. Employment protection legislation and informality: Theory and evidence from India. Available at SSRN 4204033.

Coles, Peter, Cawley, John, Levine, Phillip B, Niederle, Muriel, Roth, Alvin E, Siegfried, John J, 2010. The job market for new economists: A market design perspective. J. Econ. Perspect. 24 (4), 187–206.

Coles, Peter, Kushnir, Alexey, Niederle, Muriel, 2013. Preference signaling in matching markets. Am. Econ. J. Microecon. 5 (2), 99–134.

Crépon, Bruno, Duflo, Esther, Gurgand, Marc, Rathelot, Roland, Zamora, Philippe, 2013. Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. Q. J. Econ. 128 (2), 531–580.

Dammert, Ana C., Galdo, Jose, Galdo, Virgilio, 2015. Integrating mobile phone technologies into labor-market intermediation: A multi-treatment experimental design. IZA J. Labor Dev. 4 (1), 11.

Diamond, Peter A., 1982. Wage determination and efficiency in search equilibrium. Rev. Econ. Stud. 49 (2), 217–227.

Escudero, Verónica, Kluve, Jochen, López Mourelo, Elva, Pignatti, Clemente, 2019. Active labour market programmes in Latin America and the Caribbean: Evidence from a Meta-analysis. J. Dev. Stud. 55 (12), 2644–2661.

FICCI, 2013. Reaping India's promised demographic dividend.

Franklin, Simon, 2018. Location, search costs and youth unemployment: Experimental evidence from transport subsidies. Econ. J. 128 (614), 2353–2379.

Groh, Matthew, McKenzie, David, Shammout, Nour, Vishwanath, Tara, 2015. Testing the importance of search frictions and natching through a randomized experiment in Jordan. IZA J. Labor Econ. 4 (1), 7.

India Skills Report, 2019. Confederation of Indian industry. https://www.aicte-india.org/sites/default/files/India.pdf.

International Labour Organization, 2003. Industrial training institutes in India; the efficiency study report. URL http://temp.oitcinterfor.org/public/english/region/ampro/cinterfor/news/gasskov.pdf.

LinkedIn, 2015. Why and how people change jobs. https://business.linkedin.com/content/dam/business/talent-solutions/global/en_us/job-switchers/PDF/job-switchers-global-report-english.pdf.

LinkedIn, 2017. New report reveals the trends that will define recruiting in 2017. https://www.linkedin.com/business/talent/blog/talent-strategy/trends-that-will-define-recruiting.

Mamgain, Rajendra P., 2019. Formal Labour Market in Urban India: Job Search, Hiring Practices and Discrimination. SAGE Publications India.

McKenzie, David, 2017. How effective are active labor market policies in developing countries? A critical review of recent evidence. World Bank Res. Obs. 32 (2), 127–154.

Mortensen, Dale T., Pissarides, Christopher A., 1994. Job creation and job destruction in the theory of unemployment. Rev. Econ. Stud. 61 (3), 397–415.

Pallais, Amanda, 2014. Inefficient hiring in entry-level labor markets. Am. Econ. Rev. 104 (11), 3565–3599.

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