Matching, Management and Employment Outcomes: A Field Experiment with Firm Internships*

Girum Abebe[†], Marcel Fafchamps[‡], Michael Koelle[§] and Simon Quinn[¶]

February 2026

Abstract

Young professionals in developing economies face significant barriers to high-quality professional employment. We evaluate a novel intervention in Ethiopia that places recent graduates into established firms to shadow middle managers. Using random assignment into program participation, we find that the one-month placement significantly increases the probability of wage employment in the short run. Six years later, treated individuals see a 12% increase in average earnings, driven by a shift among the top decile of the wage distribution. We rule out signalling and social networks as primary drivers, finding instead that the effects are rooted in the accumulation of practical managerial skills and familiarity with professional organisational practices. We run the experiment using a firm-proposing deferred-acceptance algorithm to match professionals with firms. We then develop a generative model of preferences and use this for counterfactual mechanism design. Our results show that the choice of matching algorithm is pivotal: under random matching, the program would likely have yielded no significant short-run impacts. Our results demonstrate that incorporating participant preferences via matching algorithms can improve the design and efficacy of field experiments.

KEYWORDS: field experiments, matching, active labour market programmes, propensity score.

*We are grateful to Oriana Bandiera, Natalie Bau, Nick Bloom, Emily Breza, Gharad Bryan, Stefano Caria, Katherine Casey, Jonathan de Quidt, Dave Donaldson, Tore Ellingsen, Matthew Elliott, Avi Feller, Robert Gibbons, Ben Golub, Selim Gulesci, Seema Jayachandran, Terence Johnson, Max Kasy, Dorothea Kübler, Shengwu Li, Attila Lindner, David McKenzie, Eva Meyersson Milgrom, Ben Olken, Aureo de Paula, Benjamin Roth, Rafaella Sadun, Tom Schwantje, Ran Shorrer, Meredith Startz, Tavneet Suri, Adam Szeidl, Alex Teytelboym, Stefan Trautmann, Chris Udry, Gabriel Ulyssea, John Van Reenen and Chris Woodruff for useful comments, and to audiences at Bocconi University, Boğaziçi University, DIW Berlin, Harvard University, Imperial College London, the Lahore School of Economics, Lancaster University, the London School of Economics, the NBER Summer Institute, the Massachusetts Institute of Technology, Northwestern University, Queen Mary University of London, Stanford University, the Stockholm School of Economics, Trinity College Dublin, University College London, the World Bank, the University of Nottingham and the University of Oxford. We thank Gezahegn Gebremedhin, Lir Hoxhaj, Laura Lufray, Tom Schwantje, Biruk Tekle and Alemayehu Woldu for outstanding research assistance. We thank the UK Department for International Development for research funding (through the Private Enterprise Development in Low-Income Countries program). The opinions expressed and arguments employed herein are solely those of the authors and do not necessarily reflect the official views of the OECD or its member countries. Our pre-analysis plan is available at https://www.socialscienceregistry.org/trials/2776/history/49533.

[†]**International Finance Corporation**: gtefera@ifc.org
[‡]**Freeman Spogli Institute, Stanford University**: fafchamp@stanford.edu
[§]**Organisation for Economic Cooperation and Development, Paris**: michael.koelle@oecd.org
[¶]**Imperial College London**: simon.quinn@imperial.ac.uk.

1 An internship experiment for young professionals

Young adults face many barriers to enter the labour force and kick-start their careers. Even those who obtain a good education struggle to overcome search frictions (Rogerson et al., 2005; Mortensen and Pissarides, 1999), the inability to signal their skills (Spence, 1973), vague or biased expectations about wages and other job attributes (Abebe et al., 2025; Duflo et al., 2021; Conlon et al., 2023), and limited access to social and professional networks (Dustmann et al., 2015; Kramarz and Skans, 2014). Labour market exclusion of young job-seekers is particularly prevalent in developing countries (Quintini, 2014), likely reflecting greater market frictions (Donovan et al., 2023). This holds back employment, productivity and the efficient allocation of talent.

A growing number of experimental studies have tested the relevance of these mechanisms and evaluate interventions to overcome them (Carranza et al., 2022; Abebe et al., 2021, 2025; Alfonsi et al., 2020; Pallais and Sands, 2016; Dal Bó et al., 2013; Beaman and Magruder, 2012). A key omission, however, is the role of practical on-the-job experience, which is typically accumulated endogenously and has been shown to substantially shape the career paths of young adults (Blundell et al., 2016; Keane and Wolpin, 1997; Rosen, 1972; Ben-Porath, 1967). Practical experience involves skills in multiple dimensions (Lise and Postel-Vinay, 2020; Heckman and Sedlacek, 1985), and increasingly includes social skills (Deming, 2017). Cross-country evidence on life cycle wage growth shows that experience-wage profiles are only half as steep in developing countries compared to advanced economies, and even shallower in the informal sector (Abel et al., 2022), suggesting slow human capital accumulation over the course of average working lives (Lagakos et al., 2018).

We run a novel field experiment, designed to test the effects of exposing young professionals to a month of work experience shadowing managers of successful firms. In order to do so, we create a new institution to match and place highly educated young professionals inside established medium and large firms where they spend a one-month 'management experience'. Management placements are a potentially valuable way of learning and diffusing social and managerial skills, particularly in a developing country context. We act as market maker by identifying a suitable population of firms and a pool of early-stage young professionals and by matching the professionals with internships in firms. The field experiment takes place in Ethiopia, a fast-growing economy with few established private-sector firms. The experiment gives each randomly selected participant the opportunity to shadow a middle manager in their daily work, in order to see and experience first-hand the managerial realities of medium and large-scale organizations.

Host firms are representative of businesses in Addis Ababa and surroundings, and operate in different sectors, at different scales, and with different management practices.

We use two different assignment mechanisms in our experiment — each of which facilitates a different kind of causal estimation. First, we use *random assignment to select which participants enter the program* and which are assigned to the control group. This allows us to estimate the average treatment effect of being invited to the program as it was implemented. We evaluate the impact on young professionals' labour market outcomes, and on their management skills, attitudes, and practices. At the six- and 12-month mark, we find that offering a management placement to highly educated and motivated young Ethiopians increases their success in obtaining a good permanent wage job and increases their hours and earnings in wage employment. It has no average effect on their propensity to run a firm on their own, and no significant effect on the profits of firms run by this group. By the 72-month mark, we find no impact on the probability of employment – but we find large and significant impacts on wage income. Specifically, we estimate an increase in average wages of about 12% of the control group mean; we show, using a quantile-quantile plot, that this is driven by a highly significant increase in wages among the top 10% of earners.

We are able to rule out a number of mechanisms through which the management experience placement could have affected jobs outcomes. We did not provide any certification until after the (12-month) endline survey, ruling out signalling effects. We do not find any effects on changed beliefs, such as reservation wages or profits. While treated individuals have slightly larger professional networks, this likely only reflects their greater propensity to be in professional jobs: they are hired into those jobs through a formal process after responding to open calls for applications rather than informal networks or referrals, and we can confidently rule out hiring by the firms that individuals were placed in.

We do find evidence of greater accumulation of practical skills from the management experience placement, in line with the type of experience provided by the treatment. Participants were in fairly close and intense contact with the managers they shadowed, on a daily basis. After the culmination of their placement, participants report an increased confidence in their management abilities along several dimensions that are consistent with the tasks that the average participant would have completed. Those participants that ran their own firm, scored better on a standard management practices measure. Even though the programme only lasted for a month, it is unlikely that junior professionals in large organisations in the control group would have organically accumulated a similar exposure to a manager's daily activities – and if so, over a much longer timespan. In fact, anecdotal accounts tell us that more than one participant took leave or resigned from their normal job to take part in the programme.

Second, among those participants entering the program, we use a *firm-proposing deferred-acceptance algorithm to match individuals with host firms*. The controlled structure of this matching algorithm allows us to obtain propensity scores at the level of the individual-firm match; in turn, this allows us to estimate average outcomes under alternative assignment mechanisms. Specifically, we exploit three features of our research design to estimate such counterfactuals. First, matches are assigned via a controlled and fully replicable mechanism that is based only on expressed preferences in the form of rankings. Second, we control the information each participant – individuals and firms – has to form rankings. Third, the composition of peers is exogenous to individual outcomes: after each round of applications, individuals were invited to the programme in random order. Together, these three features facilitate a peer composition identification strategy.

In our specific application, we estimate a generative Bayesian classification model to characterize the determinants of individuals' ranking of potential host firms, and the rankings of individuals by firms. The results of this model are then used as a basis for integrating over many plausible counterfactual assignments, exploiting the 'equal treatment of equals' property of our controlled matching. This allows us to treat variation in matches among 'equals' as purely driven by the stochastic composition of peer environments; as such the propensity score is a sufficient statistic for the assignment mechanism (Rosenbaum and Rubin, 1983). In doing so we build on recent empirical advances from the mechanism design literature (Abdulkadiroğlu et al., 2017; Borusyak and Hull, 2023). We note that similar methods are applicable to a variety of matching experiments where matches are assigned according to a known formula. This covers in principle a broad array of interventions, such as consulting and training (Brooks et al., 2018; Bruhn et al., 2018), education (Banerjee et al., 2007), and matching in networks (Breza and Chandrasekhar, 2019).

Using this method, we find several key advantages of using the deferred acceptance algorithm rather than a random match. First – and perhaps unsurprisingly – we show that individuals are substantially more likely to report being satisfied with their match under deferred acceptance than they would be under random matching. Second, we estimate that this increases individuals' engagement with the programme. Finally, we estimate that essentially all of our estimated short-run labour market impacts can be explained by the use of the deferred acceptance algorithm: across several key outcomes, the estimated counterfactual impact of random matching is similar in magnitude and opposite in sign to our estimated Average Treatment Effects of the programme. Other results from the same exercise imply that we would have observed larger Average Treatment Effects had we used an algorithm that placed more weight upon individuals' preferences (for example, Individual-Proposing Deferred Acceptance, or Individual-Proposing Random Serial Dictatorship). This illustrates and quantifies the potential gain from using assignment mechanisms in policy experiments that are replicable and which take into account participants' preferences, and from evaluating the effects of alternative assignment mechanisms.

Our work relates to two distinct bodies of literature. First, our work lies within the experimental literature on identifying and overcoming barriers to jobs and employment in developing countries (for recent reviews, see Caria et al. (2024) and Breza and Kaur (2025)). In Ethiopia, where our intervention took place, this literature has documented – among other barriers – costly job search and applications (Caria et al., 2023; Abebe et al., 2021; Blattman and Dercon, 2018), signalling (Abebe et al., 2021), and distorted beliefs (Abebe et al., 2025). Our intervention is novel in that it provided a time-limited practical experience in close contact with a firm's management, with some similarity to internships. This sets it apart from previous research that randomly offered apprenticeships (Alfonsi et al., 2020; Hardy and McCasland, 2023; Crépon and Premand, 2025) or low-skilled jobs (Breza et al., 2021; Blattman and Dercon, 2018)—both in the type of exposure to firms and the type of participants, urban and well-educated youth.

While the literature tends to focus on disadvantaged youth from low-income backgrounds (Alfonsi et al., 2020; Carranza et al., 2022), our sample consists exclusively of young professionals with high educational attainment.[1] Our 72-month follow-up allows us to benchmark the long-term returns of such interventions against the current frontier of development research. Our documentation of a 12% earnings increase at six years extends the multi-year evidence base established by Crépon and Premand (2025) and Abebe et al. (2021), both of whom find significant earnings impacts at a four-year horizon. However, our results are mechanistically distinct: whereas the gains in Crépon and Premand (2025) are driven by productivity in self-employment, and those in Abebe et al. (2021) by overcoming initial search frictions, our results point toward a long-term transition into the upper tail of the formal wage distribution. Our findings suggest that the effects for these graduates are driven by the specific managerial skills acquired from this experience, rather than through other constraints that a match with a firm might have relaxed.

[1] One exception is Abebe et al. (2021), whose sample also consists of young jobseekers in Addis Ababa with either a university degree and college diploma, and is therefore most comparable to ours.

Second, from a methodological perspective, our paper adds to a growing body of research that applies theoretical insights from mechanism design to field experiments (Jayachandran et al., 2017; Hussam et al., 2022; Narita, 2021). While most of this line of work employs mechanism design to elicit truthful reporting or to incentivize behavior, our approach utilizes mechanism design for encouraging program take up – and, importantly, in the service of causal inference. In this respect, we build on the structural evaluation of two-sided matching markets pioneered by Bates et al. (2024), who utilize generative preference models to simulate counterfactual policy outcomes in labour markets.

Our results demonstrate that assignment by mechanism can fundamentally increase program effectiveness, providing experimental validation for concerns that 'preference-blind' matching can lead to significant labour market inefficiencies (Crépon and Pernaudet, 2024). This relates our work to the empirical literature on centralized allocation and school choice (Abdulkadiroğlu et al., 2017; Trapp et al., 2023). More generally, this research relates to several recent two-sided field experiments matching firms with job-seekers (Alfonsi et al., 2020; Brown et al., 2024; Crépon and Premand, 2025; Abebe et al., 2025; Groh et al., 2015).[2]

The paper proceeds as follows. In Section 2, we describe the experimental setting in Addis Ababa, the recruitment of young professionals and host firms, and the implementation of the management placement program . Section 3 presents the average treatment effects on short-run and long-term labor market outcomes, including our main results on the 72-month wage impacts and the exploration of skill-based mechanisms . In Section 4, we develop our framework for counterfactual mechanism design, specifying the generative model of preferences used to simulate alternative assignment algorithms. This section presents the results of these simulations, quantifying the performance of the deferred-acceptance algorithm relative to random matching and other counterfactual regimes. Section 5 concludes by discussing the implications of our findings for the design of labor market interventions and the professional socialization of high-potential youth.

[2] Alfonsi et al. (2020) use random matching conditional on region and sector; Hardy and McCasland (2023) and Brown et al. (2024) use random matching conditional on a firm being listed as acceptable; Abebe et al. (2025) use a deferred-acceptance algorithm; and Crépon and Premand (2025) and Groh et al. (2015) use manual or algorithmic matching.

2 Setting and experiment

2.1 Young professionals in Addis Ababa

We conducted a two-sided field experiment with firms and young professionals in Addis Ababa, and in the nearby towns of Adama and Bishoftu. The experiment involved a novel programme that we designed: placing young professionals in established firms, where they would shadow a manager in his or her daily activities, for a month. This modality of the placement programme resembles internships – which are known in this context for certain professional careers – and we were careful to emphasize to all parties involved that the purpose was to learn about the day-to-day activities of a manager.

The fieldwork was carried out in 2016 and 2017 by staff at the Ethiopian Development Research Institute. Both sides – young professionals and firms – were randomised to participate in the program. Young professionals and firms in the treatment groups were paired using an algorithmic mechanism described below. We evaluate the impact of this program on young professionals' labour market trajectories and outcomes, their skills, knowledge and attitudes about management and entrepreneurship, as well as potential channels that link treatment to these outcomes of interest. Given the two-sided randomisation, we can also detect any potential impacts on firms.[3]

Our experiment was designed for a highly educated population who had expressed an interest in management and entrepreneurship. We recruited with this aim in mind: we limited eligibility to young Ethiopians (aged 18 to 30, inclusive), having a minimum of technical/vocational, college or university qualifications (the 'young professionals'). We advertized using a combination of social media, college campus visits and postings on city 'job boards', using a headline message designed to attract aspiring managers and entrepreneurs. We approached firms in two ways. First, we recruited among firms in Addis Ababa that had already been part of another study (Abebe et al., 2021). Second, we updated that firm listing with new data that had become available in the interim, as well as with data provided by the respective municipal authorities in Adama and Bishoftu. We drew new random samples, weighted by industry employment shares, and subject to a minimum size and turnover threshold. Our experimental sample consists of firms that completed a baseline survey and confirmed their interest to our enumerators. We created gathered fields of these

[3] Appendix A.1 contains further details on recruitment of young professionals, benchmarking of professionals and firms, the randomization procedure for young professionals, the implementation of the DA algorithm, and a short summary of the program implementation based on a debriefing survey and administrative records we collected.

firms according to their date of completion of the baseline survey (subject to availability expressed by the firm) and randomized firms by computer within each gathered field.

2.2 Context and data

Addis Ababa is an excellent location to test this sort of programme. The Ethiopian economy has been growing steadily over the last two decades (admittedly from a very low initial level). The country is landlocked and has a large population, creating a captive market for locally produced goods. With no colonial legacy but a recent history of socialist rule, the country has few established foreign investors and suffers from a shortage of large, well-managed firms.

In line with the eligibility criteria that we specified, the experimental sample was drawn from a highly educated sub-population (Table 1). Three out of four completed at least an undergraduate college degree; the remainder possess a vocational or technical post-secondary qualification. The median schooling is 15 years. The most frequent degrees are in engineering and related STEM subjects, and in business studies.

We worked with these young professionals at a critical transition period in their lives. Half the sample graduated in the year before they entered our programme, or in the same year. At the moment of their induction session, only 25% of participants were employed in some form of wage job, and another 7% ran their own business. 80% had undertaken any job search activity in the last month, and all of those who searched were looking for a professional or managerial job. 30% had thought of starting a business, but very few had moved from having a business idea to taking actual steps. When we conduct a follow-up interview a year later, almost 70% of the control group have found a wage job, and 13% are self-employed.[4] It is in this context of rapid entry into a professional career that our intervention takes place.

And indeed, most participants are off to a very successful career that takes them to the top of the wage distribution. According to the 2021 Labour Force and Migration Survey, the average salary in Ethiopia was around ETB 4,000 per month (around USD 90 at the time). At our six-year follow-up in 2022 – when the average participant is only 30 years old – 90% of those with a wage job in the control group earn more than that, with an average (median) wage of ETB 12,000 (ETB 8,700) per month. Nine out of ten of those jobs are in professional or managerial roles and with a permanent contract, which is what most job seekers aspire

[4] The job seekers seemed to have had realistic expectations about the wages they could eventually earn – see Appendix Figure A.4 which compares reservation wages to realized wages in the sample at baseline and a year later. Expectations about profits in self-employment seem to be placed too high for at least the bottom half of the distribution (see Appendix Figure A.5).

to. By contrast, in another sample of relatively educated urban youth in Addis Ababa recruited by Abebe et al. (2021), only 25% of control participants held a permanent job at endline. The biggest factor explaining this difference is the education level: whereas all our participants had a degree or diploma, only 13% in the sample of Abebe et al. (2021) did so.

Host firms are medium to large establishments operating in a variety of sectors in the economy. Most firms are located in Addis Ababa, while around 15% are located in two mid-sized towns in central Ethiopia: Adama (about 90km from Addis Ababa) and Bishoftu (about 45km from Addis Ababa). The firms operate mainly in services (about 40%), manufacturing (about 25%), trade (about 20%) and other sectors. The median firm has 57 employees ($Q_1 = 22$, $Q_3 = 155$). The firms are nationally representative in their size distribution. Management practices of these firms are towards the bottom of the cross-country distribution reported by Bloom et al. (2012).[5] Our questionnaire embeds their survey instrument and hence management practice scores are directly comparable.

2.3 Implementation

Our randomisation yields balanced treatment and control samples. For young professionals, we applied pairwise stratified randomisation on education, age and gender. Our randomization is well balanced on all outcome variables, as we report in Appendix Table A.1. Firms were stratified on the date of their baseline survey (through the gathered-field design), but not paired. We find that firms are balanced on most variables other than management scores (Appendix Table A.2). This is due to treated firms being more likely to monitor any production performance indicators. We control for any baseline differences in the dependent variable through ANCOVA.

For logistical reasons, we conducted the program on a rolling basis, in 42 weekly batches with, on average, 40 young professionals (of whom about 20 would be assigned a placement) and 8 firms per group. We purposefully selected firms into batches according to their availability. This proved to be important to ensure the participation of firms. We invited applicants in a locally random order to an information session about the program.[6] We randomized participation in the program among the young professionals who turned up at each induction session using a physical randomization device. The Appendix provides further details. We refer to the young professionals that were assigned to participate as 'interns'.

Within each batch, we invited all interns to rank all host firms, and all host firms

[5] We show this descriptively in Appendix Figure A.3.

[6] We accepted applications for several rounds, and randomized the order within each round.

to rank all interns. Firms were given a short (anonymous) CV of interns that included gender, age, education level, field of study, and higher education institution, and work experience (both in time and by industry). Appendix Figure A.7 shows the form we used for the CV. Interns were given the following information about the firm: name of the firm, sector, approximate location, and size category of the firm. We enforced completeness and transitivity of the rankings: all interns had to rank all firms within a batch (and vice versa); ties were not allowed.

We then matched all interns and firms within the batch, using the deferred-acceptance (DA) stable matching algorithm first proposed by Gale and Shapley (1962). We describe our implementation of DA matching in Appendix A.1. We implement the 'firm-proposing' version of the DA algorithm. Firms and young professionals were notified of their respective matches; in general, participants would start their placement in the week following their induction sessions (typically on the Monday). Generally, participants were accompanied and introduced to the firm on their first day of placement by one of our field staff.

The main result from assigning placements by DA matching is that most interns and firms end up with a preferred counterpart. Appendix Figure A.1 shows that firms'rankings of young professionals positively correlate with interns' rankings: the young professionals a particular firm prefers are, on average, likewise inclined to prefer that firm. Figure 1 shows the implication of this for the matches that were generated: this figure displays the relative rank of the counterparts to which interns and firms were matched. The figure shows that around 40% of interns and 45% of firms were matched with a counterpart in their top 20%. In most batches, this corresponds to interns being assigned to their top 1 or 2 firms.[7] Overall, more than 70% of interns and firms end up matched with a counterpart in their top 40%. As a further descriptive statistic on this matching context, we implement the algorithm of McVitie and Wilson (1971), to calculate the size of the core of stable matches for each of the batches. We find that the cores are generally reasonably small: the median batch has just two stable matches (that is, the firm-proposing deferred acceptance solution and the individual-proposing deferred acceptance solution), and the range is from one to six.

We can define take-up of our program by young professionals in several ways, shown in Appendix Table A.4. Of all the 829 young professionals assigned to treatment, 788 (95%) completed the process of being matched to a firm and 588 (71%) completed at least one day at the firm. From qualitative evidence and an exit survey with treated young

[7] Firms have slightly more preferred counterparts, which would be expected since they are the proposing side in our DA algorithm (Gale and Shapley, 1962; Roth and Sotomayor, 1990).

professionals, the most common reasons for not completing the placement were (i) holding or finding a job with no possibility to take leave of absence, (ii)the location of the firm, and (ii) personal/family reasons.

2.4 Data

We collected baseline surveys with all young professionals just before randomisation; we then conducted an exit survey with treated individuals and collected administrative data on program completion. We followed up with in-person surveys, six months, 12 months, 48 months and 72 months after they completed their placement (and at equivalent moments for the control group, who were paired to treated individuals for the purpose of randomization). The 48-month survey was interrupted by COVID; we completed approximately half of this follow-up wave before the pandemic struck, and then used phone surveys with a shortened questionnaire to complete the wave. We also conducted monthly phone surveys for the first year after completing the placement, to learn about job search and employment trajectories. We analyse response rates in Appendix Section A.3. There we show that (i) response rates are high for all survey rounds (approximately 95% for the first two follow-ups, then 82% at the 72-month mark), (ii) that attrition is generally uncorrelated with treatment,[8] and (iii) that attrition cannot be predicted by any of a set of key baseline covariates.

We surveyed firms when they declared availability for the program, and again shortly after the program had ended (and we paired control units, here for the sole purpose of balancing the time of the survey).

3 Average treatment effects

3.1 Identification strategy

We begin by pooling the six-month and 12-month surveys (we show follow-up results from 48-months and 72-months, and disaggregate these rounds, in section 3.5). We use the following ANCOVA specification:

$$y_{ipt} = \beta_1 \cdot T_i + \beta_2 \cdot y_{ip0} + \delta_p + \varepsilon_{ipt}, \tag{1}$$

[8] We have a significant treatment effect only at the 12-month mark, where treated respondents are 3 percentage points less likely to attrit.

where T_i is the treatment dummy for respondent i, and y_{ipt} refers to the outcome for i in stratified pair p at follow-up time t (where we denote the baseline as $t=0$ and the 6-month and 12-month follow-ups as $t=1$ and $t=2$ respectively)[9]. Our coefficient of interest is β_1, which we interpret as the *intent-to-treat* (ITT) estimate. For each hypothesis test on the coefficient of interest, we report the usual p-value from a Wald test; we also report False Discovery Rate q-values, taken across the coefficients of interest within an outcome family (Benjamini et al., 2006). We winsorize all continuous outcome variables (within each survey wave) at the 95th percentile.

Prior to the first analysis of our experimental follow-up data, we registered a detailed pre-analysis plan.[10] We pre-specified the estimation equation, the main family of outcome variables of interest – occupation and earnings – as well as outcome families to test for learning about management from the experience provided by the placement. We further specified a few tests for other mechanisms, in particular related to job search. We augment our pre-specified tests with further analysis to contextualise and help interpret our main results, and to conduct counterfactual mechanism design.

3.2 Occupation and earnings

Table 2 reports our pre-specified main outcomes: this shows the impacts of treatment on young professionals' labour market outcomes – in particular, on occupation and on earnings.

We begin by testing for impacts on wage employment. In column 1, we find an increase in the probability of permanent employment of four percentage points; this is significant at the 5% significance level, and remains significant after correcting for multiple hypothesis testing. In column 2, we find a positive impact of three percentage points on the likelihood of having any kind of wage job, and this is close to significant: $p=0.11$.[11] In column 3, we obtain a positive estimate on the probability of having a wage job with managerial responsibilities (about 2 percentage points, relative to a control mean of 12% at follow-up). However, this is not statistically significant ($p=0.23, q=0.27$). These headline impacts at the extensive margin are broadly in line with estimates from recent search and matching experiments in urban low-income settings: Carranza and McKenzie (2024); Caria et al. (2024).

Of particular interest here are the results on hours per day in wage work, and

[9] Following De Chaisemartin and Ramirez-Cuellar (2024), we cluster at the pairwise level.

[10] The pre-analysis plan is available at `https://www.socialscienceregistry.org/trials/2776/`.

[11] This likely reflects some substitution of casual work for permanent jobs. Franklin (2018) and Abebe et al. (2021) report a similar substitution for transport subsidy interventions in Addis Ababa.

monthly wage income (where wage income is zero for those not in wage work). In column 4, we test for effects on wage work hours. We find an increase of about 0.4 hours per weekday (on a control group follow-up mean of about 4.9 hours). In column 5, we find a large and significant impact on wage income: this increases by about 300 birr on a follow-up control group mean of about 2600 birr. (At the time of the study, one US dollar was approximately 30 Ethiopian birr.) Does this increase merely reflect the extensive-margin impact described earlier, or does it also indicate an intensive-margin improvement in the conditional wage distribution? To answer this question, we use the method of Attanasio et al. (2011) to construct plausible bounds on the hourly wage and daily hours. We calculate bounds on the impact on the hourly wage of 4 birr and 74 birr, and on daily hours of 0.2 to 0.3 hours; this suggests that, in addition to improving the probability of finding permanent work, our treatment also had positive impacts on hourly wages and on hours worked conditional on holding a job.[12]

These results are a stark contrast to our results on self-employment – which we test in column 6. We find zero effects on self-employment: this is precisely estimated both at the extensive and intensive margin. The point estimate is 0.004, and is not statistically different from zero ($p = 0.77, q = 0.60$), compared to a control mean at follow-up of 0.12. This is a tight estimate of a zero effect: we can rule out a 2.5 percentage point increase with 95 percent confidence. Similarly, we estimate a zero effect on hours in self-employment (column 7), and on self-employment profits (where profit is set to zero for those not in self-employment; column 8).[13]

Are these results surprising? Before our first public presentation of results from the experiment, we elicited predictions from three distinct groups of experts (DellaVigna and Pope, 2018): international academics working on causal inference, graduate students in development economics at Oxford, and HR experts from Ethiopia. We explained the experiment, showed them the self-employment and wage-employment rates at six and twelve months of the control group, and asked for a prediction of the corresponding figures for the treatment group. The results are in Appendix Table A.5. All groups of experts were overconfident in

[12] The estimated employment effect is 0.032 and the probability of employment in the treated group is 0.672. For the hourly wage, the 10th percentile of the distribution among employed members of the control group across the first two follow-up waves is 212; the 90th percentile is 938. For the hourly wage, the bounds are calculated as $[41.4 - 0.032 \times (303.7 + 41.4)/(0.640 + 0.032)]\ /0.640 \pm (938 - 212) \times 0.032/0.672$. For daily hours, the 10th percentile of the distribution of wage work hours among employed members of the control group across the first two follow-up waves is 8 hours; the 90th percentile is 9 hours. For daily hours, the bounds are calculated as $[0.40 - 0.032 \times (4.90 + 0.40)/(0.640 + 0.032)]/0.640 \pm (9 - 8) \times 0.032/0.672$.

[13] Our results therefore indicate that, in this context, the brief management experience afforded by the programme did not provide a stepping stone for starting successful business, a mechanism which has been documented elsewhere for former managers in large firms (Muendler et al., 2025)

the effect of the program on self-employment. Experts expected a self-employment rate of 16-19% at six months and 17-29% at twelve months, compared to actual rates in the treatment group of 12% and 15% respectively. On the whole, academics made the most accurate predictions, with a forecast error of 0-4 percentage points. Academics and students were close with their predictions on wage-employment rates. Ethiopian HR experts overpredicted the effect of the program on self-employment, and underpredicted the wage-employment rates of interns, perhaps even expecting a substitution away from wage-employment. Given their small number ($n = 5$), however, HR expert group predictions have a much higher variance than the other two groups. These results mirror findings from Casey et al. (2023) that prior beliefs of experts from academia for programs with mixed and modest effects can be reasonably accurate, whereas local experts tend to expect more profound impacts from interventions.

The results in Table 2 show (i) that our treatment had a positive impact on the probability of having a wage job and (ii) that this seemed to generate better jobs – in the sense of paying a higher hourly wage and being more likely to be full-time. To understand how this result came about, we first dig further into the types of jobs that treated respondents found, and then explore potential mechanisms.

3.3 Effects on job characteristics

In Table 3, we test for treatment effects on job characteristics. Here we test for impacts on work tasks: we use dummy outcome variables for (i) whether respondents have a permanent job doing manual work, (ii) whether respondents have a permanent job doing clerical work, (iii) whether respondents have a permanent job doing professional work, and (iv) whether respondents have a permanent job doing managerial work (a memorandum item from Table 2, column 3).[14] We find a large and significant impact on the professional work category: in effect, all of the headline treatment effect on permanent work is explained by a shift into professional tasks; with some individuals also moving into managerial tasks. At the same time, there is a statistically significant substitution away from clerical and manual tasks. In Appendix Table A.6, we test for treatment effects across the distribution of firm sizes. Specifically, we form quintiles of the distribution of firm size among wage-employed respondents in the control group. We then run five separate regressions – each testing the impact of treatment on the probability of wage employment in one of those five quintiles. We find that the impact is distributed relatively uniformly across the firm-size quintiles; it is *not* the case

[14] These categories are not mutually exclusive: respondents were allowed to list multiple classes of work task for their job.

(for example) that treatment was simply helping respondents into wage jobs in small firms. If anything, respondents are somewhat more likely to move into mid-sized firms, although the effect is not statistically significant after correcting for multiple hypothesis testing.

In sum, the management experience placement helped respondents move into wage jobs over the course of the next year, reflected in higher employment probability, hours, and monthly wage earnings by about 10% each relative to the control group. Those wage jobs are more likely to be of the kind sought by our respondent group: professional, permanent, full-time roles that come at times with some managerial responsibilities and pay a higher hourly wage. We find some evidence of substitution away from casual, non-professional and more precarious wage jobs, and no effect at all on self-employment.

3.4 Mechanisms

How did these effects come about – and what should we learn from these effects about labour market constraints facing educated young professionals? Broadly, there are three main channels through which our effect could have arisen: (i) through expanding treated respondents' social networks (generating new hires either in the host firm or in other firms), (ii) through changing respondents' job-search preferences (by encouraging treated respondents to prefer wage employment over self-employment), (iii) through building skills (whether task- or sector-specific skills, or more general skills). We consider each potential channel in turn.

3.4.1 Social networks and referrals

First, we find no evidence that the treatment effects operated through a social network channel – namely, retention or hiring by the host firm and employment through job referrals (including informal search and personal referrals). This is true in terms of potential hiring at the host firm: at the 12 month follow-up survey, we find only two out of 829 interns working at their host firm. It is also true of the role of social networks more generally; when we asked young professionals at the 12-month follow-up survey to list business contacts, only one listed a contact from their host firm (whereas 60 listed contacts from their current employer).

We test this more generally and formally in Table 4. In that table, we report a series of regressions, each of which tests whether the respondent holds a permanent job that was found through a particular search method. We find that the entire impact on permanent work is explained through respondents finding jobs through open calls and with formal hiring processes: indeed, we estimate tight zero impacts on the probability of finding permanent jobs through referrals, informal search, or personal referral. In Appendix Table A.7, we complement this analysis by testing for treatment effects on respondents' own reports of finding a job through their social networks. We estimate tight zero effects on the number of potential referrers, on whether respondents list anyone who might refer them for a job, on whether they were hired after a social referral, and on whether the respondent received a job through a social contact. (Indeed, looking at the control group means, we find that this kind of informal referral behaviour is very unusual among young Ethiopian professionals generally.)

3.4.2 Job search and job preferences

Second, we test whether the treatment persuaded the young professionals to change their career aspirations – and, therefore, to tilt their search activity towards wage employment rather than self-employment. For example, respondents might have adjusted their reservation wages (either becoming more realistic or more ambitious) or their search behaviour in one or the other sector. One difficulty in testing for this mechanism lies in the fact that, even though we are able to elicit beliefs and job search behaviour, those could be different for the treatment group due to a fundamental change in job prospects rather than differences in preferences or attitudes. Our tests therefore combine outcomes on employment, job search, and stated attitudes from both the monthly phone survey data – which allows us to test for attitudes and search at a monthly frequency over the first year – and pre-specified analysis from the follow-up surveys.

We first assess respondents' beliefs about their future employment status. In Figures A.8 and A.9 of the Online Appendix, we show monthly impacts on whether respondents considered it likely (or very likely) that they would be wage employed 12 months in the future; we asked the same question about self-employment. We find a positive impact on both margins, which fades away during the year over which we observe the respondents. From the six- and 12-month follow-up surveys, and at monthly frequency from the phone survey, we can observe whether respondents follow through with any adjustment to their job search. In Figures A.10 and A.11 in the Online Appendix, we show treatment and control trajectories for (i) the probability of searching for a wage job and (ii) the probability of planning to start a business; as well as a test for the difference between treatment and control groups. We estimate zero impacts on both margins. In Appendix Figure A.12, we show positive impacts on wage employment; this reflects what we already reported from the follow-up surveys in Table 2).[15]

Similarly, in Table A.8 of the Online Appendix, we test for treatment effects on steps to start a business and on reservation profits; and in Tables A.9 and A.10 on treatment effects on steps to find a wage job and reservation wages. We find no effect on steps to start a business, and no effects overall on wage search. (We find some evidence that is statistically significant on reduced search for professional jobs; this is only individually significant.) With an effect size similar to the treatment effect on holding professional work, this likely reflects again a discontinuation of job search by those that found a job.

Finally, we find no evidence of changes to reservation wages needed to accept a job, or the minimum level of expected profits respondents would require to start a business. If anything, point estimates are positive (but not statistically significant) and therefore inconsistent with a correction of overoptimistic beliefs driving job finding (unlike in a similar context in Ethiopia studied by Abebe et al. (2025)). In sum, none of these results are consistent with a channel by which treatment tilted career aspirations and search behaviour beyond the mechanical substitution by those that successfully found wage employment.

3.4.3 Skills and attitudes about management

Finally, we test for the third hypothesised mechanism: learning about management by programme participants (and, related, diffusion of management from host firms to the interns they hosted). We approach this measurement challenge in various ways. First, we ask young professionals to self-report their *confidence* in their skills to succeed in various management tasks. Second, we ask respondents about the *relative* importance of various management practices. This would allow us to at least measure any changes in attitudes. Third, for those young professionals who run their own business at follow-up, we can measure the management practices they apply in their firms; using a survey instrument specifically calibrated to small firms in developing countries by McKenzie and Woodruff (2017).

Young professionals were remarkably confident in their management skills even before they were assigned to our program. On average, respondents at baseline state that they are either confident or very confident in their skills across 10 out of 14 areas of management. Appendix Table A.12 presents the wording of each question. In columns (1) and (2) of Appendix Table A.14, we report the effect of treatment on two summary measures of management confidence: the sum of positive responses, and a normalized

[15] Self-employment rates for treatment and control groups are indistinguishable at any point after treatment; this is in line with the regression findings (we show the equivalent graphs for self-employment in Appendix Figure A.13).

index. We find that treatment significantly increases confidence by either measure.[16] We note that we see effects on confidence only in areas that young professionals were most exposed to in host firms: planning tasks such as cost and demand estimation, or dealing with clients (Appendix Table A.13). We generally see no effects on areas that were off limits to most interns such as dealing with suppliers, and access to finance. This gives us reassurance that, despite the clear overconfidence bias in these measures, respondents do pick up some signal about the effects of the program on perceived managerial skills.

Next, we report differences in rankings across management practices. We anticipated that we would not be able to obtain an accurate measure of knowledge of best practice in management through a survey measure, given that informed respondents would have a good sense that picking the most sophisticated structured practice would be the 'right' answer. We therefore opted to elicit choices about trade-offs by asking young professionals to rank ten 'good' management practices – some more relevant for small firms, and some more relevant for large. Figure 2 shows the distribution of first-ranked practices for the treatment and control group (the sum across categories has to be equal to one). The most important practice for control respondents is separating household and business assets (18%), the least important one is frequently monitoring employee performance (3%). We find that the distributions across groups are significantly different ($p = 0.06$). Interns are more likely to put weight on practices associated with the large host firms they were placed in; in particular sales targets and employee monitoring.

Finally, we report the average effect of treatment on management practices in firms run by respondents. In columns (3)–(7) of Table A.14, we report standardized effects on overall practices, and sub-components marketing practices, record-keeping practices, and financial practices, following (McKenzie and Woodruff, 2017). We find some suggestive evidence that placement increased overall management practices. The point estimate is 0.08 standard deviations, and is significant at the 10% level, though only before correcting for FDR. Estimates on individual components of practices are highly non-significant. By definition, we only observe management practices for those respondents who run a business, so our estimates will reflect a mix of selection and learning about management. To explore this issue further, we split the sample into incumbents – who already had a business at baseline – and entrants (Appendix Table A.15). We find that all the effects are driven by incumbents, where the point estimate of treatment on overall practices is 0.16 standard

[16] This test survives correcting for testing of 16 hypotheses in the family, including the individual areas reported in Appendix Table A.13.

deviation (only weakly significant). The point estimate for entrants is zero.

In sum, these results show (i) a diffusion of the subjective importance of different management practices from the medium-to-large host firms to interns, (ii) an increase in the self-reported confidence in succeeding in practical management tasks (in line with the kinds of tasks that the average participant carried out), and (iii) suggestive evidence of improvement of management practices among the subset of participants who were self-employed before participating in the programme and continued that activity after their participation. Taken together, these results imply that practical learning – from the experience of shadowing a manager – was likely an important mechanism by which interns were better able to secure professional jobs.

3.4.4 Heterogeneous effects

To further explore the mechanisms for our effects, we estimate heterogeneity in our treatment effects. We do this using the generic machine learning method of Chernozhukov et al. (2025). We summarise our implementation of this method in Appendix A.4. Point estimates from this method indicate that the least-affected quartile had a zero treatment effect on the probability of permanent employment; this rises to a treatment effect of about 8.1 percentage points for those in the most-affected group.[17]

Of particular interest in our context are the baseline covariates that differ significantly between those in the most-affected quartile and those in the least-affected quartile. In several key respects, these covariates describe traditional proxies of labour market success: those in the most-affected quartile are significantly more likely to have a degree and to be educated in engineering; similarly, they have higher baseline scores in maths, in digit recall, in self-confidence and in the internal locus of control.

At the same time, those most affected hail from family and personal backgrounds that traditionally proxy labour market exclusion: in particular, relative to the least-affected quartile, those in the most-affected quartile have substantially less educated parents. Further, those most affected were less likely at baseline to speak English (as measured in their English comprehension score), and less likely already to be wage employed at baseline. This suggests that treatment is particularly effective for high-potential individuals who may face traditional barriers to labour-market entry. This finding highlights a key mechanism of the

[17] We cannot reject the null hypothesis that all four quartiles had equal effects; specifically, when we test equality of treatment effects for the most-affected and least-affected group, we obtain $p = 0.135$. This should not be surprising, given the sample size and the estimated effect sizes.

program: it provides high-ability graduates with the organisational familiarity and social capital often missing from their family backgrounds, allowing them to translate underlying talent into professional success. This interpretation resonates with similar findings by Abebe et al. (2026) in their study of managements styles of young professionals, and the value that employers place on them.

3.5 Long-term follow-up

We now turn to the long-term follow-up data – collected 48 months and then 72 months after assignment to treatment. Table 5 includes both follow-up surveys, and further disaggregates our short-term results by survey wave.

Our 48-month survey was launched just before the COVID pandemic. We completed about a third of our intended interviews in person before the pandemic; we then switched to phone surveys, with a smaller questionnaire, during the pandemic; Table 5 disaggregates between these two parts of the 48-month follow-up survey. This pandemic disruption limits what we can learn from the 48-month survey. Nonetheless, several suggestive insights emerge. First, despite the reduced sample size, we see suggestive evidence that the treatment effects on employment were sustained into the pre-pandemic period. The magnitude of treatment effects on wage employment outcomes is similar to the magnitudes at the 6- and 12-month marks, as reported earlier in Table 2. Second, by the 48-month mark, we find suggestive evidence that our treatment helped respondents to substitute into wage employment (with a positive coefficient of 5 percentage points) and out of self-employment (a negative coefficient of 5 percentage points).[18] This is consistent with the high overall rate of employment among our pre-COVID sample at the 48-month mark: in the control group, for example, 92% of respondents are employed (70% in wage employment and 22% self-employed).

Two years later, in 2022/23, the employment rate in the control group had returned to pre-pandemic levels. At this 72-month mark, we no longer find any treatment effect on employment status. However, we do find striking evidence of a treatment effect on wage income. Specifically, we find an effect of 947 birr – about 12% of the control group mean – significant at the 5% level. (We note that – relative to the control group mean – this effect size is of similar magnitude to the short-term effect.) In Figure 3, we explore this result further – with a quantile-quantile plot of the distribution of monthly wages at the 72-month

[18] For completeness, we show coefficients on profit income in column 8. None of these impacts are positive; we note that reported coefficients are large and negative, but we note that – given the skewness of the data – the standard errors here are also very large.

point. This plot shows that the increase in average earnings is driven by a shift at the top of the distribution: the distributions are remarkably close until about the 90th percentile, but then diverge markedly. (We use quantile regressions to test for distributional equality at specific points, generating p-values using randomisation inference. For the 90th, 95th and 99th percentiles respectively, we obtain $p = 0.132$, $p = 0.013$ and $p = 0.015$.)

Taken together, our long-term follow-up results suggest that, beyond short-term impacts on employment, wages and job quality in the first year, the management experience placement programme had impacts on the longer-term career trajectories of participants: it increased their likelihood of ultimately building their career as employees in organisations, and it increased their chances of obtaining a high-paying position in those organisations.

3.6 Testing for effects on host firms

We also test for effects on host firms, exploiting the fact that firms were also randomized into participation. We have firm data from a baseline survey, and from a follow-up survey shortly after the conclusion of the program. We again estimate an ANCOVA specification. We cluster standard errors at the level of firms' gathered field.

Hosting a young professional for a short period of time is unlikely to transform how medium and large firms do their business. Nevertheless, the profile of interns is potentially different from that of firms' usual employees, and we anticipated that hosting such interns might change some aspects of firm behaviour (Loiacono and Silva-Vargas, 2024). We therefore test for effects on a few selected and pre-specified HR and management outcomes, a few weeks after the placement. We find no effects on firms' advertising, hiring or general management practices (Appendix Tables A.16, A.17, and A.18). We obtain a significantly positive point estimate on separations, which implies that treated firms had an additional 3 workers separations (A.17). We find no effect on hiring. However, we cannot rule out that this reflects the mechanical effect from our program, which placed on average 2 interns into host firms. If we assume that all firms include their program interns in the calculation of separations, the coefficient reduces to 1.1 ($p = 0.29$).[19]

[19] We conduct a number of additional tests that were pre-specified in our pre-analysis plan, but which we do not discuss in the main text. These can be found in the Online Appendix.

4 Peer composition and counterfactual mechanism design

Like most forms of human capital acquisition, the placement experience likely led to highly heterogeneous skills accumulation, part of which might depend on the specificities of the match. We explained earlier that our experiment used a firm-proposing deferred-acceptance algorithm to match individuals with host firms. We decided to use this algorithm – rather than a random assignment – in order to improve the quality of matches. What, then, was the consequence of using the firm-proposing deferred-acceptance algorithm? Did the use of this algorithm meaningfully improve upon the use of a random assignment – and would some other algorithm have fared better? These questions are important for understanding the mechanisms behind our experimental results, and central for thinking about potential future scale-up of such assignment programmes. In this section, we build a generative model of interns' and firms' preferences; we then show how the results of this model allow counter-factual mechanism design through numerical integration.

To begin, we note that – because our assignment takes into account firms' and interns' preferences – the matching of interns to firms is likely to reflect match-specific potential outcomes. One common way that such selection can be introduced is through differential compliance (Alfonsi et al., 2020). This is particularly relevant in two-sided labour market experiments, where compliance requires both the worker and the firm to agree to the placement; unsurprisingly, low compliance often characterises two-sided experiments (Alfonsi et al., 2020; Groh et al., 2015). More generally, in the presence of heterogeneous treatment effects, the average treatment effects (ATE) documented in the last section are inevitably a function of the particular way the programme was assigned (Heckman, 2005). We stress that this result is not limited to the algorithmic matching that we use in our current study; it is a feature of any non-commodity treatment, whether assigned explicitly (e.g. random matching to treatment providers) or implicitly (e.g. purposeful matching based on location, availability, or other observable characteristics).

The advantage of having used a controlled assignment mechanism in our setting is that it allows us to determine what assignment would look like under a range of counterfactuals. Three features of our research design are key to this. First, within the randomly chosen treatment group, matches are assigned via a replicable mechanism that takes as its sole input the expressed preferences of both participants and firms. Second, we control the information each side has access to in order to express their preferences in form of a rank order; this prevents unobservables from contaminating expressed preferences. Third, the composition of peers is exogenous to individual outcomes: after each round of applications, individuals were invited to an induction session in random order. Together, these features facilitate a peer composition identification strategy for counterfactual mechanism design.

The intuition of this approach is straightforward. Suppose, as a thought experiment, that two identical interns are assigned to two different batches – but that one faces tougher competition from peers. This generates quasi-exogenous variation in the realised match. Specifically, we can exploit the replicable nature of our assignment mechanism (Borusyak and Hull, 2023) to obtain a propensity score for each individual-firm match, by treating the composition of peers as a random variable that can be integrated out by simulation (Sørensen, 2007; Hortaçsu and McAdams, 2010; Hortaçsu and Kastl, 2012; Cassola et al., 2013).

4.1 Notation

We index interns by $i \in \{1,...,N\}$ and firms by $f \in \{1,...,F\}$. The $N \times F$ matrix $\boldsymbol{Y}$ denotes potential outcomes – such that element Y_{if} is the potential outcome of variable Y if intern i is assigned to firm f. (We denote row i of $\boldsymbol{Y}$ as $\boldsymbol{Y}_i$.) The $N \times F$ matrix $\boldsymbol{m}$ describes realised matching – such that $m_{if} = 1$ indicates that intern i was matched to firm f. The $N \times F$ matrix $\boldsymbol{r}^I$ records interns' ranking of firms; we denote its i-th row as the $1 \times F$ vector $\boldsymbol{r}^I_i$. Symmetrically, $\boldsymbol{r}^F$ is the $F \times N$ matrix recording firms' ranking of interns (where elements r^I_{if} and r^F_{fi} are set to zero – without loss of generality – if i and f were in different batches). Denote by $\boldsymbol{x}_{if}$ the K-dimensional vector of observable covariates describing intern i and firm $f \in \{1,...,F\}$; these include the firm characteristics provided to interns before the ranking exercises and the intern characteristics listed on the anonymised and standardised CVs showed to firms, as well as any interactions between these characteristics.[20] For intern i, we then stack all $\boldsymbol{x}_{if}$ into the $K \times F$ matrix $\boldsymbol{x}_i$. Denote by $\boldsymbol{z}_{-i}$ the matrix that stacks $\boldsymbol{x}_{if}$ for all individuals $j \neq i$. In order to treat the composition of peers as a random variable to be integrated out, we treat $\boldsymbol{z}_{-i}$ as the realisation of a random variable $\boldsymbol{Z}_{-i}$; similarly, $\boldsymbol{m}$ is the realisation of the random variable $\boldsymbol{M}$. We treat $\boldsymbol{r}^I$ and $\boldsymbol{r}^F$ respectively as realisations of the random functions $\boldsymbol{R}^I$ and $\boldsymbol{R}^F$. We describe our matching algorithm – firm-proposing deferred acceptance – using the function ψ, which takes as its inputs the matrices $\boldsymbol{r}^I$ and $\boldsymbol{r}^F$ and returns the matrix of realised matches:

$$\boldsymbol{m} = \psi\left(\boldsymbol{r}^I, \boldsymbol{r}^F\right). \tag{2}$$

[20] In practice, we use just one interaction: namely, a dummy for whether the intern and firm are in the same general part of Addis; we allow this to matter for interns' rankings of firms. Note that, as we explain shortly, for any firm f not in the same batch as intern i, $\boldsymbol{x}_{if}$ is redundant; however, we define $\boldsymbol{x}_{if}$ over the full set $f \in \{1,...,F\}$ for notational simplicity.

In this setting, the average structural function (ASF) is the average outcome over all $N \times F$ possible matches:

$$\frac{1}{N \cdot F} \sum_{i=1}^{N} \sum_{f=1}^{F} Y_{if}. \tag{3}$$

In contrast, the observed average is the average outcome over N realised matches (where we use the notation Y_{if^*} to refer to the realised outcome for intern i):

$$\frac{1}{N} \sum_{i=1}^{N} Y_{if^*} \equiv \frac{1}{N} \sum_{i=1}^{N} \sum_{f=1}^{F} m_{if} \cdot Y_{if}. \tag{4}$$

We anticipate that the ASF will differ from the realised average – even in large samples – since matches were formed non-randomly (so that $\mathbb{E}\left(M_{if} \mid \boldsymbol{Y}\right) \neq \frac{1}{F}$).

4.2 Generating propensity scores

With this notation in hand, we can define the propensity of i matching with each f conditional on the characteristics of i and the characteristics of each firm in i's batch, and on i's ranking of those firms:

$$p_{if} \equiv \Pr(M_{if} = 1 \mid \boldsymbol{x}_i, \boldsymbol{r}_i^I) = \mathbb{E}_{\{\boldsymbol{Z}_{-i}, \boldsymbol{R}_{-i}^I, \boldsymbol{R}^F\}} \Big[\psi([\boldsymbol{r}_i^I, \boldsymbol{R}_{-i}^I(\boldsymbol{Z}_{-i})], \boldsymbol{R}^F([\boldsymbol{x}_i, \boldsymbol{Z}_{-i}])) \Big]_{if}. \tag{5}$$

That is, even though our assignment mechanism is deterministic, we can think of the characteristics of each intern's peers as a shifter for the propensity of each intern being matched to each firm in the batch. We note that due to locally random assignment of interns to batches, this shifter is exogenous to i – their characteristics, their preferences, and also their potential outcomes; a fact we exploit shortly. By taking the expectation over these characteristics ($\boldsymbol{Z}_{-i}$) and over the consequent rankings expressed by firms ($\boldsymbol{R}^F$) and by other interns ($\boldsymbol{R}_{-i}^I$), we can calculate a propensity score for each intern-firm match.[21]

We take this expectation using numerical integration. This requires a generative model for the random variable $\boldsymbol{Z}_{-i}$ and for the random functions $\boldsymbol{R}^F$ and $\boldsymbol{R}^I$. In principle, several different methods could be used to construct these objects; none of our earlier

[21] By contrast, we do hold fixed the characteristics of firms in i's batch, for two reasons. First, firms were not assigned to batches randomly but purposefully, based on their availability to host interns. Second, holding fixed i's rankings – as i's preferences might correlate with their potential outcomes – requires holding fixed the set of firms.

reasoning depends upon the particular estimation method used to characterise participants' preferences. Different methods have been used in similar empirical work to estimate primitives that serve as an input to simulating counterfactual outcomes under known assignment mechanisms, such as auctions (Hortaçsu and McAdams, 2010; Hortaçsu and Kastl, 2012; Cassola et al., 2013).[22] We draw from the marginal distribution of $\boldsymbol{Z}_{-i}$ by using a discrete non-parametric density estimate over intern characteristics; specifically, apply a weak Dirichlet prior across all possible categorical realisations, and update using the observed characteristics of interns in the given batch, and in batches occurring at similar times. For the random functions $\boldsymbol{R}^F$ and $\boldsymbol{R}^I$, we need a generative model of interns' preferences over firms; symmetrically, we need a generative model of firms' preferences over interns. For this we use a flexible ranking model (specifically, a discrete finite mixture version of a Plackett-Luce model), estimated using a Hamiltonian Markov Chain with Maximum Likelihood estimation for post-processing (where we choose the number of types using the BIC). We discuss this model in detail in Appendix A.5, and briefly describe model results here.

4.3 Estimated preferences

We use three types to characterise intern preferences over firms. The first type comprises nearly half of interns (about 44%); this type values firms that are geographically proximate, relatively large, and outside the manufacturing sector. Their observable characteristics indicate a higher likelihood of prior wage employment and a background in the social sciences. The second type, comprising about 38% of interns, instead displays strong preferences for manufacturing and trading firms; members of this type tend to be younger and have weaker prior employment histories. The smallest group (about 16%) places substantial weight on construction firms and is disproportionately composed of engineering students.

We use four types to characterise firm preferences over interns. The largest group comprises about 38% of firms, and displays almost no systematic structure in its ranking of interns; its choices are close to random. These firms tend to be smaller and score low on the management-quality index. The second type – with about 25% – shows clear preferences for business-major interns and is disproportionately composed of large firms with high management scores. A third type – about 23% – favours both engineering and business

[22] Sørensen (2007) estimates preferences in a two-sided market with a structural model that imposes both an empirical structure for valuations, as well as a structure given by a Gale-Shapley algorithm about how valuations translate into matches. Our application differs in that the Gale-Shapley algorithm is not a *model* but rather embodies the true assignment mechanism.

interns and is concentrated in construction and manufacturing—sectors that draw more heavily on technical skill. Finally, about 14% mirror the second type in preferring business students, but represents another segment of large, high-management firms.

4.4 From propensity scores to counterfactual outcomes

Following equation 4, the expected value of the realised outcome for intern i matched with firm f – conditional on that intern's characteristics and expressed rankings – is:

$$\mathbb{E}\Big(M_{if} \cdot Y_{if} \,|\, \boldsymbol{x}_i, \boldsymbol{r}_i^I, \boldsymbol{Y}_i\Big) = Y_{if} \cdot \Pr(M_{if} = 1 \,|\, \boldsymbol{x}_i, \boldsymbol{r}_i^I) = Y_{if} \cdot p_{if}. \tag{6}$$

The first equality in equation 6 follows from an important property of controlled assignment mechanisms: the 'equal treatment of equals'. Consider two interns $i \neq j$ having the same characteristics, ranking and peer composition (that is, $(\boldsymbol{x}_i, \boldsymbol{r}_i^I, \boldsymbol{Z}_{-i}) = (\boldsymbol{x}_j, \boldsymbol{r}_j^I, \boldsymbol{Z}_{-j})$). Then – because the matching function ψ is anonymous, and because we controlled the information set available to both interns and firms – those two interns face the identical probability distribution over potential firms f. This symmetry ensures that the propensity score is a sufficient statistic for the assignment mechanism, allowing us to treat the variation in matches among 'equals' as purely driven by the stochasticity of peer environments (Rosenbaum and Rubin, 1983; Abdulkadiroğlu et al., 2017; Borusyak and Hull, 2023).

Rearranging,

$$Y_{if} = \mathbb{E}\left(M_{if} \cdot \frac{Y_{if}}{p_{if}} \,\middle|\, \boldsymbol{x}_i, \boldsymbol{r}_i^I, \boldsymbol{Y}_i\right). \tag{7}$$

Following Horvitz and Thompson (1952) (and applying the Law of Iterated Expectations) an unbiased estimator of equation 7 is given by the sample analogue; this implies the following unbiased estimator of the ASF:

$$\frac{1}{N} \cdot \sum_{i=1}^{N} \frac{Y_{if^*}}{p_{if^*}}. \tag{8}$$

The estimator of the average structural function is therefore the inverse propensity weighting estimator – where the propensity is defined at the level of the intern-firm match.[23] Extend-

[23] Appendix Figure A.17 shows the distribution of p_{if}. We have $p_{if} > 0$ for all observed matches, and we winsorize at $p_{if} = 0.01$ (Khan and Tamer, 2010).

ing the logic, suppose that we obtain propensity scores by re-running the integration in equation 5 for some other mechanism (for example, intern-proposing deferred acceptance); denote those alternative propensity scores as $\tilde{p}_{if}$. Then the expected counter-factual average outcome – under that alternative mechanism – is:

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{f=1}^{F}\tilde{p}_{if}\cdot Y_{if} \tag{9}$$

with estimator:

$$\frac{1}{N}\cdot\sum_{i=1}^{N}Y_{if^*}\cdot\frac{\tilde{p}_{if^*}}{p_{if^*}}. \tag{10}$$

That is, we estimate the outcome under some counter-factual mechanism by reweighting the observed outcomes based on the ratio of propensity scores under that alternative mechanism to the propensity scores under the actual mechanism used.[24]

Before using our propensity scores for estimation, we run two specification checks. First, we compare the probability of a given firm-intern match with the constructed propensity p_{if}; that is, we check at the dyadic intern-firm level whether the mean assignment probability in the data corresponds closely to the simulated assignment probability. For example, of those firm-intern dyads having $p_{if}=0.4$, we expect about 40% actually to have been matched. To check this, we run a non-parametric regression of m_{if} on p_{if}.[25] We show, in Appendix Figure A.17, that the relationship indeed follows the 45-degree line very closely.

Second, we use the propensity scores to test for baseline balance between matched interns and firms. This operates as a placebo test: by construction, we know that the true ASF must be zero for outcomes measured *prior* to the management placement. In Appendix Table A.24, we test baseline balance over a series of outcomes, for key firm characteristics (namely, firm management quality, firm size and whether firms are externally managed).[26] We find very good balance on Average Structural Functions across all regressions.

[24] This is conceptually equivalent to importance sampling, where the weights account for the change in the assignment probability measure. See, for example, Heckman and Vytlacil (1999) and Dal Bó et al. (2021) – and, for applied counterfactual mechanism design, Calsamiglia et al. (2020), Bobba et al. (2021), Larroucau and Rios (2022) and Dahlstrand (2022).

[25] Specifically, we do this using a kernel regression, with a log transformation of the simulated probability of match; we use a bandwidth of 0.2 in that log space.

[26] Specifically, for intern i matched with firm f in batch b, we implement equation 8 by regressing $y_{ifb}=\beta\cdot X_f+\mu_b+\varepsilon_{ifb}$, weighting by $1/p_{if}$ where p_{if} is the propensity score for the match. For firm management quality, X_f is a dummy for whether f had a MOPS management score at or above the median at baseline. For firm size, X_f is a dummy for whether the number of employees in firm f was at or above the median at baseline. For external management, X_f is a dummy for whether firm f was externally managed at baseline.

4.5 Results

We estimate counter-factual impacts on three classes of outcomes: (i) interns' ranking of their realised match; (ii) features of the internship experience; and (iii) labour market outcomes.

Figure 4 shows counterfactual empirical CDFs for interns' ranking of their realised matches.[27] The figure shows the proportion of interns who would be matched with their first-preferred firm, then the proportion matched with a firm in their top two, then the proportion matched with a firm in their top three, and so on. As noted in our earlier discussion of Figure 1, our chosen algorithm – firm-proposing DA – performed well in placing interns with preferred firms. For example, Figure 4 illustrates that about 38% of interns were matched to their top-ranked firm, and about 73% of interns were matched to a firm in their top three. This is substantially higher than under random matching or firm-proposing Random Serial Dictatorship – each of which would have placed only about 14% of interns with their top-ranked firm, and about 40% with a firm in the interns' top three. Indeed, it is striking how similar the counter-factual outcomes are between random matching and firm-proposing Random Serial Dictatorship; this is consistent with our observation earlier that a large share of firms ranked interns in essentially a random manner. Figure 4 also illustrates the improvement in realised ranking quality for interns under alternative algorithms that prioritise their preferences (Roth and Sotomayor, 1990). For example, we estimate that intern-proposing deferred acceptance would place about 53% of interns with their best match, and about 87% with a match in their top three; under intern-proposing RSD, those figures increase respectively to about 66% and 89%.

Does this make a difference for interns' perceptions and participation? In Table 6, we estimate counter-factual impacts on key aspects of the internship experience. In columns 1 and 2, we consider participation – both at the extensive margin (whether the intern completed a full week: column 1) and the intensive margin (total weeks completed: column 2). In both columns, we estimate higher participation under algorithms giving more weight to intern preferences (namely, individual-proposing DA and individual-proposing RSD), and lower participation under algorithms giving less weight to intern preferences (namely, firm-proposing RSD and random matching). The effect magnitudes, however, are not particularly large: for example, 70% of interns completed at least one week under firm-proposing DA, and we estimate that 66% of interns would have done so under either

[27] To do this, we apply equation 10 sequentially for different values of $X \in \{1,...,10\}$, where Y_i takes a series of dummy variables of the form "intern i is matched with a firm having ranking less than or equal to X". In practice, we implement equation 10 by running a series of regressions of outcomes Y_i on a dummy, using appropriate weights; in doing so, we include demeaned batch dummies.

firm-proposing RSD or random matching.

In columns 3 and 4 of Table 6, we test impacts on interns' self-reported satisfaction – both prior to the internship (but after the match was formed) and after the internship. In column 3, we find quite large impacts on the probability of intern satisfaction: for example, we estimate a reduction of about eight percentage points had we used firm-proposing RSD or random matching (compared to 70% satisfaction under firm-proposing DA); we estimate that interns would have been about four percentage points more likely to be satisfied under individual-proposing DA, and about six percentage points more likely under individual-proposing RSD. Interestingly, these differences disappear when we ask about satisfaction after the internship (column 4). Finally, we test effects on time allocation during the internship. We find no effect on the proportion of time spent with management (column 5), but we estimate a small increase in the proportion of time spent idle under random matching (column 6).

Finally, we estimate counter-factual impacts on labour outcomes. In Table 7, we consider short-run outcomes: specifically, as in Table 2, we pool the six- and 12-month follow-up data, using the same outcomes as in that earlier table. Our primary comparison is between the actual mechanism used – firm-proposing deferred acceptance – and a random match. We estimate quite large negative counter-factual impacts of random matching compared to the actual mechanism used or any of the three other mechanisms considered. In particular, we estimate a reduction of three percentage points in the probability of permanent work and the probability of any wage work, and a reduction in wage work hours of about 0.27 per day; each of these coefficients is similar in magnitude – but opposite in sign – to the average treatment effects reported in Table 7. This implies that – had we used a random matching algorithm rather than firm-proposing deferred acceptance – we likely would not have found any short-run average treatment effect. In Appendix Table A.19, we repeat the exercise, using the 72-month follow-up wave. We find no differential impact across mechanisms, on any of the outcomes considered.

5 Conclusions

This paper provides evidence on the returns to professional internships for high-potential youth in a developing economy. In the short run, we find that a time-limited managerial shadowing experience leads to significant improvements in employment status and indices of job quality, as treated individuals transition rapidly into the formal wage sector. These initial gains are particularly pronounced for participants from socioeconomically disadvantaged backgrounds – suggesting that the programme may be serving to substitute for inherited professional networks and cultural capital. Over time, this initial divergence in career trajectories consolidates into a lasting impact on earnings: we document a 12% increase in average income that persists 72 months after the intervention. This long-run effect is driven by a marked sorting of treated individuals into the upper tail of the earnings distribution.

A central finding of our study is that these employment transitions are critically dependent on the matching architecture used for programme assignment. Our generative model reveals that a random assignment mechanism – a standard benchmark for many field experiments – would likely have yielded zero average treatment effects. This illustrates that, in labour markets for young professionals – where preferences over specialized roles and firm cultures are highly heterogeneous – the choice of assignment mechanism is not merely a logistical detail but a necessary condition for programme efficacy.

Further, we document a substantial 'optimization gap' between different preference-aware algorithms. While the firm-proposing deferred-acceptance (DA) mechanism implemented in the experiment was effective, our counterfactual analysis shows that intern-proposing assignment rules – such as intern-proposing DA or an intern-proposing Random Serial Dictator (RSD) mechanism – would likely have resulted in larger employment effects. This suggests that for high-educated cohorts, the frontier of design for active labour market programmes should shift toward participant-centred matching rules that prioritize the job-seeker's private information and preferences about match quality.

One very plausible mechanism underlying these gains is one of professional socialisation. In a companion follow-up study using a subsample of these respondents (Abebe et al., 2026), we take participants into a studio to record responses to a series of management vignettes. The results in that paper suggest that our internship treatment caused a shift to a 'rule-based' management style: a bundle of traits emphasizing formal procedures and organizational interests. There, we show that this effect is driven by respondents whose parents did not complete primary school. This provides separate supporting evidence for the suggestion that the internship functioned as a conduit for 'cultural capital' – allowing talented individuals to navigate and succeed within formal firm hierarchies.

Finally, our results highlight a distinct pathway for private-sector human capital accumulation in emerging markets. While recent evidence from Uganda suggests that intensive business training can drive growth by creating new high-productivity micro-enterprises (Chioda et al., 2026), our participants followed a different trajectory – integrating into existing firm hierarchies. This aligns with the framework in Muendler et al. (2025), which suggests that managerial talent may sort differently depending on the prevailing firm size distribution. In our context, the 'missing middle' of management in established Ethiopian firms appears to absorb high-potential youth who are then able to navigate formal organizational structures. Taken together, our results suggest that – for high-educated cohorts – interventions should move beyond simple search assistance, focusing instead on the deliberate matching and professional socialization required to support the expansion of formal-sector organizations.

References

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017). Research design meets market design: Using centralized assignment for impact evaluation. *Econometrica 85*(5), 1373–1432.

Abebe, G., A. S. Caria, and E. Ortiz-Ospina (2021, June). The selection of talent: Experimental and structural evidence from ethiopia. *American Economic Review 111*(6), 1757–1806.

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2021). Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City. *The Review of Economic Studies 88*(3), 1279–1310.

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, S. Quinn, and F. Shilpi (2025). Matching frictions and distorted beliefs: Evidence from a job fair experiment. *The Economic Journal*, ueaf026.

Abebe, G., M. Fafchamps, M. Koelle, S. Quinn, and T. Schwantje (2026). Management style under the spotlight: Evidence from studio recordings. *CEPR Discussion Paper DP21043*.

Abel, M., E. Carranza, K. Geronimo, and M. E. Ortega (2022). Can temporary wage incentives increase formal employment? experimental evidence from mexico. IZA Discussion Paper 15740, IZA – Institute of Labor Economics.

Alfonsi, L., O. Bandiera, V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2020). Tackling youth unemployment: Evidence from a labor market experiment in uganda. *Econometrica 88*(6), 2369–2414.

Attanasio, O., A. Kugler, and C. Meghir (2011). Subsidizing vocational training for disadvantaged youth in colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics 3*(3), 188–220.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics 122*(3), 1235–1264.

Bates, M., M. Dinerstein, A. C. Johnston, and I. Sorkin (2024). Teacher labor market policy and the theory of the second best. *Journal of Political Economy 132*(4), 1120–1165.

Beaman, L. and J. Magruder (2012). Who Gets the Job Referral? Evidence from a Social Networks Experiment. *American Economic Review 102*(7), 3574–93.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*(4), 352–365.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive Linear Step-up Procedures that Control the False Discovery Rate. *Biometrika 93*(3), 491–507.

Blattman, C. and S. Dercon (2018). The impacts of industrial and entrepreneurial work on income and health: Experimental evidence from ethiopia. *American Economic Journal: Applied Economics 10*(3), 1–38.

Bloom, N., H. Schweiger, and J. Van Reenen (2012). The land that lean manufacturing forgot? management practices in transition countries 1. *Economics of Transition 20*(4), 593–635.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). Female labor supply, human capital, and welfare reform. *Econometrica 84*(5), 1705–1753.

Bobba, M., T. Ederer, G. León-Ciliotta, C. A. Neilson, and M. G. Nieddu (2021, July). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in peru. (29068).

Borusyak, K. and P. Hull (2023). Nonrandom exposure to exogenous shocks. *Econometrica 91*(6), 2155–2185.

Breza, E. and A. G. Chandrasekhar (2019). Social networks, reputation, and commitment: Evidence from a savings monitors experiment. *Econometrica 87*(1), 175–216.

Breza, E. and S. Kaur (2025, June). Labor markets in developing countries. Working Paper 33908, National Bureau of Economic Research.

Breza, E., S. Kaur, and Y. Shamdasani (2021). Labor rationing. *American Economic Review 111*(10), 3184–3224.

Brooks, W., K. Donovan, and T. R. Johnson (2018). Mentors or teachers? microenterprise training in kenya. *American Economic Journal: Applied Economics 10*(4), 196–221.

Brown, G., M. Hardy, I. Mbiti, J. McCasland, and I. Salcher (2024). Can Financial Incentives to Firms Improve Apprenticeship Training? Experimental Evidence from Ghana. *American Economic Review: Insights 6*(1), 120–136.

Bruhn, M., D. Karlan, and A. Schoar (2018). The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico. *Journal of Political Economy 126*(2), 635–687.

Calsamiglia, C., C. Fu, and M. Güell (2020). Structural Estimation of a Model of School Choices: The Boston Mechanism versus Its Alternatives. *Journal of Political Economy 128*(2), 642–680.

Caria, S., S. Franklin, and M. Witte (2023). Searching with friends. *Journal of Labor Economics 41*(4), 887–922.

Caria, S., K. Orkin, A. Andrew, R. Garlick, R. Heath, and N. Singh (2024, February). Barriers to search and hiring in urban labour markets. Literature Review 1, VoxDevLit.

Carranza, E., R. Garlick, K. Orkin, and N. Rankin (2022, November). Job search and hiring with limited information about workseekers' skills. *American Economic Review 112*(11), 3547–83.

Carranza, E. and D. McKenzie (2024). Job training and job search assistance policies in developing countries. *Journal of Economic Perspectives 38*(1), 221–44.

Casey, K., R. Glennerster, E. Miguel, and M. Voors (2023). Skill versus voice in local development. *The Review of Economics and Statistics 105*(2), 311–326.

Cassola, N., A. Hortaçsu, and J. Kastl (2013). The 2007 Subprime Market Crisis Through the Lens of European Central Bank Auctions for Short-Term Funds. *Econometrica 81*(4), 1309–1345.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2025). Fisher–schultz lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. *Econometrica 93*(4), 1121–1164.

Chioda, L., P. Gertler, D. Contreras-Loya, and D. R. Carney (2026). Making entrepreneurs: Long term returns to training youth in business skills. *Working Paper* (34637).

Conlon, J. J., L. Pilossoph, M. Wiswall, and B. Zafar (2023). Labor market search with imperfect information and learning. *Journal of Political Economy 131*(12), 3421–3475.

Crépon, B. and S. Pernaudet (2024). Designing labor market recommender systems: How to improve human-based search. *Working Paper*.

Crépon, B. and P. Premand (2025). Direct and indirect effects of subsidized dual apprenticeships. *The Review of Economic Studies 92*(5), 2979–3023.

Dahlstrand, A. (2022). Defying Distance? The Provision of Services in the Digital Age. *Centre for Economic Performance Discussion Paper* (1889).

Dal Bó, E., F. Finan, N. Y. Li, and L. Schechter (2021). Information technology and government decentralization: Experimental evidence from paraguay. *Econometrica 89*(2), 677–701.

Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics 128*(3), 1169–1218.

De Chaisemartin, C. and J. Ramirez-Cuellar (2024). At what level should one cluster standard errors in paired and small-strata experiments? *American Economic Journal: Applied Economics 16*(1), 193–212.

DellaVigna, S. and D. Pope (2018). Predicting experimental results: Who knows what? *Journal of Political Economy 126*(6), 2410–2456.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Donovan, K., W. J. Lu, and T. Schoellman (2023, 05). Labor market dynamics and development. *The Quarterly Journal of Economics 138*(4), 2287–2325.

Duflo, E., P. Dupas, and M. Kremer (2021, June). The impact of free secondary education: Experimental evidence from ghana. *NBER Working Paper* (28937).

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2015). Referral-based job search networks. *The Review of Economic Studies 83*(2), 514–546.

Franklin, S. (2018). Location, search costs and youth unemployment: experimental evidence from transport subsidies. *The Economic Journal 128*(614), 2353–2379.

Gale, D. and L. S. Shapley (1962). College admissions and the stability of marriage. *The American Mathematical Monthly 69*(1), 9–15.

Groh, M., D. McKenzie, N. Shammout, and T. Vishwanath (2015). Testing the Importance of Search Frictions and Matching Through a Randomized Experiment in Jordan. *IZA Journal of Labor Economics 4*(1), 1–20.

Hardy, M. and J. McCasland (2023). Are small firms labor constrained? experimental evidence from ghana. *American Economic Journal: Applied Economics 15*(2), 253–84.

Heckman, James J.and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica 73*(3), 669–738.

Heckman, J. J. and G. Sedlacek (1985). Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. *Journal of Political Economy 93*(6), 1077–1125.

Heckman, J. J. and E. J. Vytlacil (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*(8), 4730–4734.

Hortaçsu, A. and J. Kastl (2012). Valuing dealers' informational advantage: A study of canadian treasury auctions. *Econometrica 80*(6), 2511–2542.

Hortaçsu, A. and J. Kastl (2012, November). Valuing dealers' informational advantage: A study of canadian treasury auctions. *Econometrica 80*(6), 2511–2542.

Hortaçsu, A. and D. McAdams (2010). Mechanism choice and strategic bidding in divisible good auctions: An empirical analysis of the turkish treasury auction market. *Journal of Political Economy 118*(5), 833–865.

Horvitz, D. and D. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*(260), 663–685.

Hussam, R. N., N. Rigol, and B. N. Roth (2022). Targeting high ability entrepreneurs using community information: Mechanism design in the field. *American Economic Review 112*(3), 861–898.

Jayachandran, S., J. de Laat, E. F. Lambin, C. Y. Stanton, R. Audy, and N. E. Thomas (2017). Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science 357*(6348), 267–273.

Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Khan, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*(6), 2021–2042.

Kramarz, F. and O. N. Skans (2014, 01). When strong ties are strong: Networks and youth labour market entry. *The Review of Economic Studies 81*(3), 1164–1200.

Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2018). Life cycle wage growth across countries. *Journal of Political Economy 126*(2), 797–849.

Larroucau, T. and I. Rios (2022). Dynamic college admissions and the determinants of students' college retention. *Cowles Foundation Discussion Paper* (2346).

Lise, J. and F. Postel-Vinay (2020, August). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review 110*(8), 2328–76.

Loiacono, F. and M. Silva-Vargas (2024). Matching with the right attitude: The effect of matching firms with refugee workers. (WP-295). Available at SSRN: https://ssrn.com/abstract=4975573.

McKenzie, D. and C. Woodruff (2017). Business practices in small firms in developing countries. *Mangement Science 63*(3), 2773–3145.

McVitie, D. G. and L. B. Wilson (1971). The stable marriage problem. *Communications of the ACM 14*(7), 486–490.

Mortensen, D. T. and C. A. Pissarides (1999). New developments in models of search in the labor market. *Handbook of Labor Economics 3*, 2567–2627.

Muendler, M.-A., J. E. Rauch, and S. M. Koyama (2025, August). Which Individuals Create Jobs? Managerial Talent and Occupational Skills. Working Paper 34158, National Bureau of Economic Research.

Muendler, M.-A., J. E. Rauch, and F. Tintelnot (2025). Which individuals create jobs? managerial talent and occupational skills. *Working Paper* (34158).

Narita, Y. (2021). Toward an ethical experiment. *Econometrica 89*(3), 1005–1036.

Pallais, A. and E. G. Sands (2016). Why the referential treatment? evidence from field experiments on referrals. *Journal of Political Economy 124*(6), 1793–1828.

Quintini, G. (2014, May). Skills at work: How skills and their use matter in the labour market. Technical Report 158, Organisation for Economic Co-operation and Development.

Rogerson, R., R. Shimer, and R. Wright (2005). Search-theoretic models of the labor market: A survey. *Journal of Economic Literature 43*(4), 959–988.

Rosen, S. (1972). Learning and experience in the labor market. *Journal of Human Resources*, 326–342.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

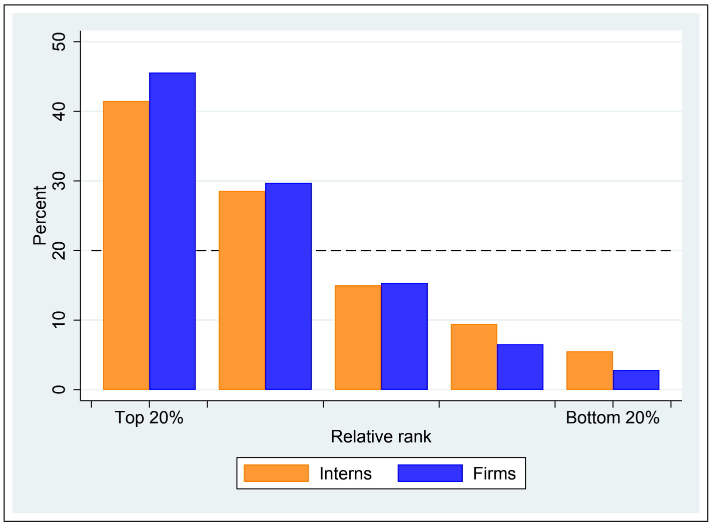
Roth, A. E. and M. A. O. Sotomayor (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monographs. Cambridge University Press.

Sørensen, M. (2007). How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital. *The Journal of Finance 62*(6), 2725–2762.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics 87*(3), 355–374.

Trapp, A. C., A. Teytelboym, A. Martinello, T. Andersson, and N. Ahani (2023). Placement optimization in refugee resettlement. *Management Science 69*(6), 3311–3333.

Figure 1: **Deferred Acceptance algorithm: Summary of assignment**



Note: This figure reports the performance of the firm-proposing DA algorithm in terms of allocating preferred counterparts to interns and firms. We report, for each quintile of the normalised rank, the fraction of firms and interns that are assigned a counterpart they rank within this quintile. Random assignment would result in bins of 20% each on average, denoted by the dashed line.

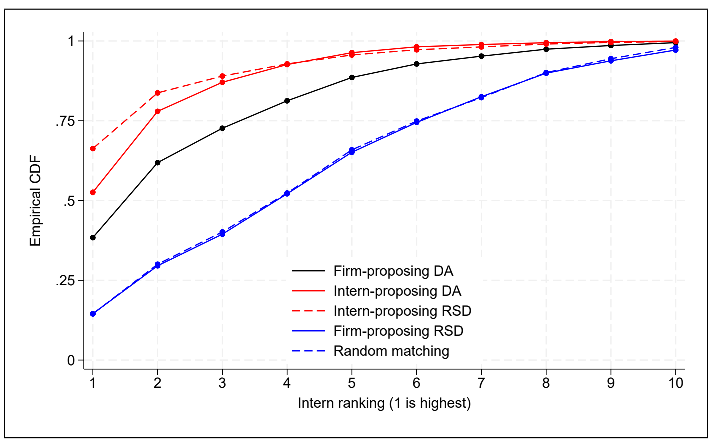
Figure 2: **Perception of most important management practice**

Note: This figure reports the distribution of which management practice interns perceive as the most important one, by treatment status. A Kolmogorov-Smirnoff test rejects distributional equality between treated and control with $p = 0.06$.

Figure 3: **Quantile-Quantile plot of wage earnings at the 72-month follow-up**

Note: This graph shows a quantile-quantile plot of wage earnings at the 72-month follow-up. It labels points at the 50th, 90th, 95th, 96th, 97th, 98th and 99th quantiles. (Data is winsorised at the 99th percentile on a wave-by-wave basis; respondents without a wage job are coded as having zero wage income.) We use quantile regressions to test for distributional equality at specific points, generating p-values using randomisation inference. We obtain $p = 0.132$ when testing equality at the 90th quantile, $p = 0.013$ when testing equality at the 95th percentile and $p = 0.015$ when testing equality at the 99th percentile. Appendix Figure A.6 shows the equivalent plot at baseline (where the quantile-quantile points sit very close to the 45-degree line).

Figure 4: **Counter-factual mechanism design: Ranking assignments**



Note: The figure shows the proportion of interns who would be matched with their first-preferred firm, then the proportion matched with a firm in their top two, then the proportion matched with a firm in their top three, and so on. The black line ('Firm-proposing DA') shows the actual algorithm used; the other lines show counter-factuals. To do this, we apply equation 10 sequentially for different values of $X \in \{1,...,10\}$, where Y_i takes a series of dummy variables of the form "intern i is matched with a firm having ranking less than or equal to X". In practice, we implement equation 10 by running a series of regressions of outcomes Y_i on a dummy, using appropriate weights; in doing so, we include demeaned batch dummies.

Table 1: **Description of young professionals**

	N	Mean	SD	p25	p50	p75
Male	1637	0.76	0.43	1	1	1
University degree	1637	0.77	0.42	1	1	1
Years of schooling	1576	15.76	1.76	15	16	17
Years since graduation	1632	2.57	2.38	1	1	4
Field of study: STEM	1626	0.58	0.49	0	1	1
Field of study: business	1626	0.21	0.41	0	0	0
Born in Addis Ababa	1635	0.27	0.45	0	0	1
Son/daughter of household head	1635	0.33	0.47	0	0	1
Household head or spouse	1635	0.37	0.48	0	0	1
Wage employed	1637	0.25	0.43	0	0	1
Monthly wage if employed	405	3657	2615	2000	3000	4600
Self-employed	1636	0.07	0.26	0	0	0
Monthly profit if self-employed	86	11914	23623	2000	5000	14000
Has searched for wage job	1636	0.80	0.40	1	1	1
Has thought of starting a business	1637	0.28	0.45	0	0	1

Notes: This table reports descriptive information on key demographic and employment variables from the baseline survey. All monetary variables are in nominal Ethiopian birr.

Table 2: **Main short-run outcomes on employment**

Outcome:	(1) Permanent work	(2) Any wage work	(3) Managerial work	(4) Wage work hours	(5) Wage income	(6) Self-employed	(7) Self-employment hours	(8) Profit income
Dummy: Treated	0.04	0.03	0.02	0.40	298.06	0.00	-0.01	-72.26
	(0.02)	(0.02)	(0.01)	(0.17)	(115.80)	(0.01)	(0.09)	(172.07)
	[0.04]**	[0.11]	[0.23]	[0.02]**	[0.01]**	[0.77]	[0.91]	[0.67]
	{0.08}*	{0.15}	{0.27}	{0.07}*	{0.07}*	{0.60}	{0.60}	{0.60}
Control mean (follow-up)	0.51	0.64	0.12	4.90	2590.12	0.12	0.70	967.18
Control mean (baseline)	0.19	0.26	0.04	1.74	861.71	0.07	0.34	338.35
Observations	3,121	3,121	3,121	3,121	3,105	3,110	3,121	3,077

Note: In this table we report the *intent-to-treat* estimates of the internship on primary employment outcomes. These are obtained by least-squares estimation of equation 1. Below each coefficient, we report a standard error in parenthesis, a p-value in brackets, and a q-value in curly braces. Standard errors allow for clustering at the level of the randomisation pair. q-values are obtained using the sharpened procedure of (Benjamini et al., 2006). We denote significance using $*$ for 10%, $**$ for 5% and $***$ for 1%.

Table 3: **Effects on formal work tasks**

Outcome:	(1) Manual work	(2) Clerical work	(3) Professional work	(4) Managerial work
Dummy: Treated	-0.01	-0.01	0.05	0.02
	(0.00)	(0.01)	(0.02)	(0.01)
	[0.04]**	[0.04]**	[0.01]***	[0.17]
	{0.05}**	{0.05}**	{0.03}**	{0.06}*
Control mean (follow-up)	0.02	0.03	0.47	0.10
Control mean (baseline)	0.00	0.01	0.17	0.03
Observations	3,120	3,121	3,121	3,121

Note: In this table we report the *intent-to-treat* estimates of the internship on primary employment outcomes. These are obtained by least-squares estimation of equation 1. Below each coefficient, we report a standard error in parenthesis, a p-value in brackets, and a q-value in curly braces. Standard errors allow for clustering at the level of the randomisation pair. q-values are obtained using the sharpened procedure of (Benjamini et al., 2006). We denote significance using $*$ for 10%, $**$ for 5% and $***$ for 1%.

Table 4: **Effects on hiring processes for formal jobs**

Outcome:	(1) Found: open call	(2) Found: job board	(3) Found: referral	(4) Hired: formal process	(5) Hired: informally	(6) Hired: personal referral
Dummy: Treated	0.04	0.00	0.00	0.04	0.00	0.00
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
	[0.04]**	[0.98]	[0.93]	[0.04]**	[0.98]	[0.79]
	{0.14}	{1.00}	{1.00}	{0.14}	{1.00}	{1.00}
Control mean (follow-up)	0.21	0.15	0.15	0.20	0.15	0.15
Control mean (baseline)	0.06	0.07	0.04	0.05	0.07	0.06
Observations	3,121	3,121	3,121	3,121	3,121	3,121

Note: In this table we report the *intent-to-treat* estimates of the internship on primary employment outcomes. These are obtained by least-squares estimation of equation 1. Below each coefficient, we report a standard error in parenthesis, a p-value in brackets, and a q-value in curly braces. Standard errors allow for clustering at the level of the randomisation pair. q-values are obtained using the sharpened procedure of (Benjamini et al., 2006). We denote significance using $*$ for 10%, $**$ for 5% and $***$ for 1%.

Table 5: **Main outcomes on employment: Disaggregated by wave**

Outcome:	(1) Permanent work	(2) Any wage work	(3) Managerial work	(4) Wage work hours	(5) Wage income	(6) Self-emp.	(7) Self-emp. hours	(8) Profit income
Treatment × 6-month	0.03	0.03	0.01	0.37*	223.16*	0.01	0.03	14.64
	(0.02)	(0.02)	(0.01)	(0.20)	(132.20)	(0.01)	(0.11)	(223.65)
Treatment × 12-month	0.06**	0.04*	0.02	0.39**	280.90*	0.00	-0.06	-451.20*
	(0.02)	(0.02)	(0.02)	(0.19)	(144.80)	(0.02)	(0.12)	(256.59)
Treatment × 48-month (pre-Covid)	0.04	0.05	-0.00	0.48	132.77	-0.05	-0.38	-1098.22*
	(0.04)	(0.04)	(0.03)	(0.32)	(478.34)	(0.03)	(0.27)	(638.03)
Treatment × 48-month (during Covid)	-0.00	0.01	0.01	.	126.83	0.03	.	.
	(0.03)	(0.03)	(0.02)		(282.64)	(0.02)		
Treatment × 72-month	-0.01	0.03	0.00	0.25	947.07**	0.02	-0.13	-914.03
	(0.03)	(0.02)	(0.02)	(0.20)	(466.58)	(0.02)	(0.19)	(1113.84)
Control mean: baseline	0.19	0.26	0.04	1.74	861.71	0.07	0.34	338.35
Control mean: 6 months	0.46	0.59	0.07	4.44	2193.79	0.10	0.61	671.10
Control mean: 12 months	0.56	0.69	0.16	5.36	2994.27	0.13	0.80	1264.43
Control mean: 48 months pre-Covid	0.63	0.70	0.13	5.30	5862.74	0.22	1.60	2591.01
Control mean: 48 months post-Covid	0.59	0.64	0.09	.	4169.68	0.13	.	.
Control mean: 72 months	0.61	0.70	0.10	5.24	7453.54	0.25	1.73	6839.04
Equality test: 6 vs 12 months	0.19	0.67	0.57	0.93	0.64	0.67	0.48	0.06
Equality test: 12 vs 48 months	0.59	0.75	0.44	0.77	0.75	0.13	0.24	0.30
Equality test: 12 vs 72 months	0.03	0.69	0.33	0.58	0.14	0.58	0.73	0.67
Total observations	5,922	5,922	5,922	5,032	5,862	5,909	5,032	4,942
Observations baseline	0	0	0	0	0	0	0	0
Observations pre-Covid wave	551	551	551	551	549	550	551	547

Note: In this table we report the *intent-to-treat* estimates of the internship on primary employment outcomes. These are obtained by least-squares estimation of equation 1 (interacting treatment with dummies for follow-up wave). Below each coefficient, we report a standard error in parenthesis.

Table 6: **Counter-factual mechanism design: Internship experience**

		(1) Any week completed	(2) Weeks completed	(3) Satisfied beforehand	(4) Satisfied afterwards	(5) Time (%) with management	(6) Time (%) spent idle
Actual mechanism:	Firm-proposing DA	0.70***	2.74***	0.70***	0.69***	58.00***	18.59***
		(0.02)	(0.06)	(0.02)	(0.02)	(1.38)	(0.95)
Difference to alternative:	Individual-proposing DA	0.01	0.04	0.04***	0.00	-0.11	-0.05
		(0.01)	(0.03)	(0.01)	(0.01)	(0.56)	(0.40)
Difference to alternative:	Firm-proposing RSD	-0.04*	-0.13*	-0.08***	0.01	1.12	1.18
		(0.02)	(0.07)	(0.02)	(0.02)	(1.33)	(1.16)
Difference to alternative:	Individual-proposing RSD	0.02*	0.08*	0.06***	0.02*	0.14	-0.14
		(0.01)	(0.04)	(0.01)	(0.01)	(1.10)	(0.76)
Difference to alternative:	Random matching	-0.04*	-0.13*	-0.08***	0.01	0.80	2.38**
		(0.02)	(0.07)	(0.02)	(0.02)	(1.32)	(1.20)
Observations		779	779	723	698	582	563
p-value (mechanisms equivalent)		0.07	0.14	0.00	0.36	0.82	0.01

Note: This table shows the estimated outcomes under counter-factual mechanisms, in comparison to the actual mechanism used. To do this, we implement equation 10 by running a series of regressions of outcomes Y_i on a dummy, using appropriate weights; in doing so, we include demeaned batch dummies.

Table 7: **Counter-factual mechanism design: Short-run labour outcomes**

		(1) Permanent work	(2) Any wage work	(3) Managerial work	(4) Wage work hours	(5) Wage income	(6) Self-employed	(7) Self-employment hours	(8) Profit income
Actual mechanism:	Firm-proposing DA	0.55***	0.67***	0.14***	5.22***	2888.20***	0.13***	0.77***	1013.99***
		(0.02)	(0.01)	(0.01)	(0.12)	(93.01)	(0.01)	(0.07)	(117.63)
Difference to alternative:	Individual-proposing DA	0.00	0.00	0.00	0.02	-23.54	-0.00	-0.04	-18.18
		(0.01)	(0.01)	(0.00)	(0.05)	(40.65)	(0.00)	(0.03)	(47.78)
Difference to alternative:	Firm-proposing RSD	-0.02	-0.02	-0.01	-0.20	-22.16	0.01	0.10	-27.66
		(0.02)	(0.01)	(0.01)	(0.12)	(102.90)	(0.01)	(0.09)	(117.41)
Difference to alternative:	Individual-proposing RSD	-0.00	-0.01	-0.00	-0.07	-61.56	-0.01	-0.10**	-90.96
		(0.01)	(0.01)	(0.01)	(0.09)	(70.38)	(0.01)	(0.05)	(72.80)
Difference to alternative:	Random matching	-0.03*	-0.03**	-0.01	-0.27**	-122.85	0.01	0.15	-22.62
		(0.02)	(0.02)	(0.01)	(0.12)	(92.03)	(0.01)	(0.10)	(104.43)
Observations		1,492	1,492	1,492	1,492	1,486	1,486	1,492	1,470
p-value (mechanisms equivalent)		0.24	0.14	0.34	0.09	0.01	0.36	0.08	0.62
Average Treatment Effect		0.04	0.03	0.02	0.40	298.06	0.00	-0.01	-72.26
		(0.02)	(0.02)	(0.01)	(0.17)	(115.80)	(0.01)	(0.09)	(172.07)

Note: This table shows the estimated outcomes under counter-factual mechanisms, in comparison to the actual mechanism used. To do this, we implement equation 10 by running a series of regressions of outcomes Y_i on a dummy, using appropriate weights; in doing so, we include demeaned batch dummies.