Hidden Discrimination in Frictional Labor Markets

Elisa Macchi* and Claude Raisaro‡

Version: January 9, 2026.

Abstract

We offer 921 Ugandan employers from three male-dominated sectors opportunities to hire male and female trainees under business-as-usual conditions and two randomly assigned monitoring regimes: audits targeting worker misbehavior/stealing and audits targeting workplace safety/harassment. These employers face monitoring constraints and perceive women as more trustworthy—the most valued yet hardest-to-find worker trait— but worry about women's harassment risk. Under business-as-usual, we observe a modest hiring gender gap, aligned with a stated unmet demand for female workers. Behavior audits reduce the demand for women workers and increase the hiring gender gap by 63%, especially among employers with stronger stated preferences for hiring women. Safety audits increase the demand for women and close the gap. Using randomized variation in technical skill signals, we show that these gaps are inconsistent with belief-based discrimination based on technical skills; moreover, employers' beliefs are also systematically inaccurate. However, estimates of bias vary widely depending on which attributes are considered productive.

*Brown University (email: elisa_macchi@brown.edu)

†Geneva Graduate Institute (email: claude.raisaro@graduateinstitute.ch)

‡We thank Siwan Anderson, Abhijit Banerjee, Aditi Bhowmick, Laura Boudreau, Lorenzo Casaburi, Peter Hull, Supreet Kaur, Rocco Macchiavello, Rohini Pande, Bobby Paksad-Hurson, Michael Peters, Jon Roth, Frank Schilbach, David Yanagizawa-Drott, Roberto Weber and workshop participants at BREAD, Brown, CEPR, KU Leuven, MIT Behavioral Lunch, NBER, NEUDC, and Zurich Behavioral Lab Meeting for helpful discussion and comments. Josh Bwiira and Carlotta Riva provided outstanding field management and research assistance. The experiments were approved by Mildmay Uganda (0408-2023) and the Uganda National Council for Science and Technology (SS2574ES), and pre-registered at the AER Registry (AEARCTR-0013698).

1 Introduction

How do labor-market frictions shape discrimination in hiring? Standard theories emphasize two channels. By limiting competition, frictions allow prejudice to persist (Becker, 1957), and by limiting information flows, they generate statistical discrimination (Phelps, 1972; Arrow, 1973; Bohren et al., 2023). Under both mechanisms, reducing frictions is expected to improve efficiency and attenuate discrimination.

This intuition is incomplete when workers' skills are multidimensional. Frictions shape not only information and competition, but also which attributes are productive. For example, monitoring constraints may raise the value of trustworthiness. When productivity depends on multiple skills, discrimination can operate along several dimensions — especially when such attributes are, accurately or stereotypically, associated with identity. As a result, changes in frictions or in the production function can alter the extent of hiring disparities and relaxing friction can also amplify disparities, revealing hidden discrimination.

This paper studies how labor-market frictions and the technologies used to address them shape discrimination in hiring when worker productivity is multidimensional. We focus on women's hiring in three male-dominated sectors in Uganda, a highly frictional labor market with high search costs, limited monitoring capacity, and strong social norms (Breza and Kaur, 2025). Despite relatively high female labor force participation, common in sub-Saharan Africa (64.9% in SSA and 76.5% in Uganda; ILO, 2025), occupational segregation by gender remains stark, with most sectors nearly single-gendered. This segregation has important implications for the allocation of talent and economic growth (Hsieh et al., 2019; Ashraf et al., 2025), as well as for gender earnings gaps (Alfonsi et al., 2024).

In a field experiment, we measure demand for female workers by offering 921 employers the opportunity to hire male or female workers under business-as-usual conditions and experimentally introduce gender-neutral technologies to reduce either monitoring frictions or safety/harassment concerns. We find substantial unmet demand for female workers, which varies systematically with the frictions firms face. These patterns align with employers' perceptions of gendered comparative advantages in relevant non-technical skills. We show that the extent of hiring disparities by gender starkly depend on the frictions at play, and a key result is that, contrary to the standard framework, reducing monitoring constraints can increase the extent of discrimination against women. Finally, we discuss the implications of such multidimensional skills for distinguishing statistical discrimination from bias.

We provide novel descriptive evidence on how the frictions firms face shape a multidimensional valuation of workers and the gendered perception of skills. Employers in mechanics, welding, and carpentry value a range of traits; however, in the presence of monitoring constraints typical of low-income settings, 90.4% identify "good behavior, honesty, and not stealing" as the hardest-to-find and monitor attribute among the five most important traits.[1] Women are perceived as systematically more trustworthy, while men are perceived as more technically skilled. Consistent with these beliefs, 84.9% of employers report a preference for hiring more women and cite limited supply as the main barrier. At the same time, employers express concerns about women's safety and harassment at the firm.

To measure demand for female workers in an incentive-compatible setting, we address supply constraints by offering employers the opportunity to hire male or female trainees from vocational training institutes in Kampala. Using an incentivized resume rating design (Kessler et al., 2019), we construct hypothetical profiles by cross-randomizing administrative data from VTI students and stratifying by gender and ability. 906 employers

[1] The five most important traits, in order of mention, are: good behavior/no stealing, effort, skills, learning interest, and cooperation.

evaluate 24 randomly selected profiles each, yielding 20,725 profile evaluations. Based on these revealed preferences, we provide tailored trainee referrals. The main outcome is the employer's decision to hire a candidate on probation.

We randomize employers to one of three conditions that subtly vary the monitoring support provided: (i) Pure Control (PC), with no audit visits, which measures the relative demand for female workers under business as usual; (ii) Monitoring–Behavior (MB), with unannounced audits aimed at deterring theft, dishonesty, and disrespect by new workers, which tests whether relaxing monitoring frictions affects the hiring gender gap; and (iii) Monitoring–Safety (MS), with unannounced audits focused on guaranteeing that new workers are safe, treated with respect, and not harassed, which tests whether alleviating safety/harassment concerns affects the hiring gender gap.[2] Because both interventions involve audit visits, employers may expect each to affect multiple dimensions of worker conduct, which could blur the distinction between monitoring and safety channels. Such symmetric spillovers would attenuate treatment differences and thus work against us finding differential effects on the demand for female workers. A stronger concern is that audit visits may generate general scrutiny effects that mechanically reduce employers' willingness to hire women under either regime. The MS arm is therefore central for identification, as it allows us to assess the presence and magnitude of such direct scrutiny effects.

Are employers interested in hiring women? Under business-as-usual conditions (PC), when provided the opportunity to hire trained women on probation, employers select women in 42.2% of cases on average. The hiring gender gap is 10.3 percentage points, conditional on observables, relative to an average selection rate for men of 52.2%. This relatively mild gap is comparable to evidence from similar sectors in Uganda (Alfonsi et

[2]The audits mimic support commonly provided by Vocational Training Institutes (VTIs), where staff visit firms during trainees' internships. We label the MS treatment "Safety" to emphasize its gender-neutral framing.

al., 2024) and smaller than in other contexts.[3] Moreover, the gap narrows significantly among employers with stronger stated preferences for hiring women, disappearing entirely in the top quintile—employers who report wanting 4 to 7 women per 10 workers. This evidence suggests an actual demand for female workers in these male-dominated sectors.

In the second part of the paper, we examine how labor market frictions affect demand for women and the hiring gender gap. We test whether these effects operate through multidimensional, gendered skills. Specifically, we study how two gender-neutral technologies—designed to address monitoring and safety frictions—affect the relative demand for female versus male workers compared to the status quo.

First, we test for the effect of behavior audits on demand for women and the hiring gender gap. Employers receiving support to monitor trainee behavior and prevent theft show a larger gender gap than Pure Control. Under MB, the gap is 16.8 percentage points—a 63% increase relative to business as usual. This increase is driven by employers with stronger stated preferences for workforce gender mix. Among employers with some preference for hiring women, the gap doubles; among those above the median, it more than triples. Relaxing monitoring constraints thus reveals latent hiring gender gaps, particularly among employers who appeared more gender-neutral at baseline.

Next, we examine the effect of safety audits (MS). We test whether MS increases the relative demand for women by alleviating concerns about harassment from coworkers or customers. Safety audits could also deter hiring by introducing scrutiny of the employer. We find support for the safety channel over the scrutiny channel. Among employers assigned to safety audits, the gender gap shrinks to 1.7 percentage points, a six-fold reduction relative to business as usual.

[3]For comparison, our estimate corresponds to a 19.7% gap; Buchmann et al. (2023) documents a gender gap of 44.7% in a male-stereotypical occupation in Bangladesh (no subsidy, no transport condition).

Together, these findings support our proposed mechanism. Technologies that address market frictions change the relative value of worker attributes. Because these attributes carry gendered connotations, they affect hiring discrimination. Demand for trust drives demand for women; when monitoring constraints are relaxed, fewer women are hired relative to men. Conversely, when harassment concerns—more binding for women in a gender-traditional environment—are addressed through safety audits, the hiring gender gap is substantially reduced. The lack of support for the scrutiny channel strengthens this interpretation: the widening gender gap under behavior audits reflects reduced demand for trust-related traits, not a response to external scrutiny.

Using randomized variation in skill signals across experimental profiles, we show that gender gaps under business-as-usual conditions and across treatments are inconsistent with statistical discrimination based on technical skills alone. This implies that the hidden gap we uncover by providing monitoring support cannot be explained by statistical discrimination based on technical skills and would therefore be taken as a measure of bias. We further show that estimates of bias, however, vary widely—from 1.9% to 16.5%—depending on which attributes are considered productive: technical skills only, trust-related traits, or safety concerns. Finally, we document that employers' beliefs about skills are systematically inaccurate, further complicating the inference of bias from observed gaps.

Our paper contributes to two strands of literature. First, to a literature studying discrimination in the labor market (Becker, 1957; Phelps, 1972; Goldin and Rouse, 2000; Bertrand and Mullainathan, 2004; Kline et al., 2022). Recent work has emphasized the role of the "system" in shaping discrimination, primarily by studying how discrimination accumulates across multiple stages of evaluation (Bohren et al., 2019, 2025). Our contribution complements this literature by highlighting a different mechanism. We show that changes in technology and market frictions (the "system") can amplify or reduce discrimination at a given hiring stage. In this sense, the choice of production technology is itself a determinant of discrimination.

These contributions stem from studying discrimination while modeling workers as multidimensional, a perspective increasingly emphasized in the literature (Deming, 2017; Woessmann, 2024; Deming and Silliman, 2025). Two additional insights arise when minority candidates are perceived as positively selected along at least one dimension. First, the absence of outcome gaps does not imply the absence of discrimination. This conclusion echoes Bohren et al. (2025), but for a different reason: observed equality in outcomes can reflect the net effect of offsetting forces—positive (statistical) discrimination and negative bias operating along different dimensions. Second, estimating bias from residual variation after controlling for technical-skill signals is problematic, as this will underestimate bias.

Second, we contribute to the literature on female labor force participation in low-income countries. While a large body of work has examined supply-side constraints to women's employment (see Jayachandran, 2015 for a review), evidence on demand-side factors remains comparatively limited (Gentile et al., 2023), particularly for the African context. Existing studies have largely examined the determinants of female labor force participation (Fernández and Fogli, 2009; Bernhardt et al., 2018) or discrimination within sectors where women are already employed (e.g., Brown, 2023; Macchiavello et al., 2020), rather than occupational segregation —despite the role for the gender wage gap (Goldstein et al., 2019). Our results speak directly to this gap. We show that there is substantial demand for trained female workers in male-dominated sectors, and that this demand is shaped by labor market frictions and employers' perceptions of women's non-technical skills. In doing so, we provide relevant evidence to the open question of whether skill training can facilitate women's entry into underrepresented occupations (Heath et al., 2024).

With a similar focus, Buchmann et al. (2023) study employer demand for female workers in a male-stereotypical occupation in Bangladesh and show that paternalistic beliefs about workplace safety can lead to gatekeeping that limits women's entry. Our results complement this evidence by showing that different frictions can have opposing effects on the gender gap, because they imply different perceived comparative advantages or disadvantages for women. Unlike other work on harassment in the workplace (Boudreau et al., 2023; Sharma, 2023), we do not assess whether our safety intervention identifies or reduces harassment; yet, our results suggest that safety audits can nonetheless increase the relative demand for female workers.

2 Setting and Sample Characteristics

2.1 Setting

Our focus is Uganda, where—like in many low-income settings—labor markets are characterized by substantial frictions, including turnover, absenteeism, difficulties in finding skilled workers, and information and monitoring constraints (Breza and Kaur, 2025). The production technology evolves endogenously to these frictions. For example, firms commonly respond to monitoring constraints through piece-rate pay, which makes effort contractible (Foster and Rosenzweig, 1994), and through reliance on social networks for hiring (Beaman and Magruder, 2012; Heath, 2018).

Female labor force participation in sub-Saharan Africa is relatively high, averaging 64.9% in 2024 (ILO, 2025). Uganda lies at the upper end of this distribution, with 76.5% of women participating in the labor force.[4] Yet despite high participation, occupational segregation is extreme—a pattern documented even in richer African contexts like South Africa, and not explained by worker observables such as education (Gradín, 2021). While the literature has identified drivers of the variation in female labor force participation across countries (Fernández and Fogli, 2009; Bernhardt et al., 2018), the determinants of occupational segregation remain less understood. Partly, network based hiring reinforces occupational segregation by limiting women's access to male-dominated sectors (Beaman et al., 2018). Beyond structural constraints such as network-based hiring, gender norms and beliefs likely play a role. Developing countries tend to have especially strong norms promoting clear occupational divisions of labor by gender (Jayachandran, 2021), and are a determinant of low female labor force participation in South Asia and the Middle East (Bursztyn et al., 2020; Agte and Bernhardt, 2024).

Occupational segregation is a key driver of wage inequality and talent misallocation (Hsieh et al., 2019; Goldstein et al., 2019; Ashraf et al., 2025). This motivates policy efforts to increase women's presence in male-dominated sectors. A common intervention involves vocational training, but impacts are often mixed and short-lived. This pattern suggests that training alone is unlikely to succeed without complementary interventions addressing norms, safety, or other demand-side constraints (Bandiera et al., 2022).

To investigate the role of these barriers, we focus on three manufacturing sectors—mechanics, welding, and carpentry—where women hold less than 3% of jobs (Alfonsi et al., 2020). Given the stark segregation described above, studying barriers to women's entry requires focusing on such male-dominated fields.

[4]In fact, Alfonsi et al. (2024) study the gender employment effects of COVID-19 among skilled Ugandan workers and, using a high-frequency panel dataset, show that the pandemic created a gender employment gap of 20 percentage points across sectors, which was previously nonexistent.

2.2 Sample Selection

Our study population is employers from motor mechanics, carpentry, and welding firms in the Greater Kampala area. Our full sample consists of 921 employers, one per firm: 318 from motor mechanics, 300 from carpentry, and 303 from welding (see Table 1).[5]

We selected firms in the neighborhoods of the greater Kampala area where these sectors are clustered, as shown in Appendix Figure A1. We conducted a listing of all firms in these neighborhoods and interviewed all those who consented to participate, conditional on meeting the following criteria: (i) working in small or medium-sized firms (fewer than 100 employees); (ii) aged 18 or older; (iii) expressing interest in hiring a trainee within nine months; and (iv) proficiency in either English or Luganda.

In our sample, 84.9% of respondents are business owners, while 15.1% hold managerial positions. The firms are well established, employing a mean of 8.3 workers (median 5). They serve approximately 11.8 clients per day, have operated for 10.7 years, and report mean monthly profits of USD 286 on revenues of USD 718.[6] In line with the figures reported above, the firms in our sample are overwhelmingly male-dominated. Fewer than 14% of firms employ any women at all. 96.4% of employers in our sample are men, and only 2.26% of all workers are women (173 out of 7,658 employees).

We also collaborate with seven major vocational training institutes (VTIs) in Kampala to collect data on trainees' backgrounds and gain access to trainees seeking employment. Our primary VTI partner is the Smart Girls Foundation, a nonprofit founded in 2012 that delivers nationally recognized vocational training through its "Girls With Tools"

[5]We initially focused on a single sector, motor mechanics. In the preregistration, we noted that we might expand to additional sectors for power and external validity, depending on the number of mechanics we were able to interview. After the first wave, we decided to expand the study to carpentry and welding in the second wave. The preregistration was amended before expanding data collection and includes a document that summarizes the motivations; details are summarized in Appendix A1.

[6]Uganda's 2024 GDP per capita was USD 1,023.

program. The organization specializes in preparing women for male-dominated sectors—including automotive mechanics, electrical installation, welding, and carpentry—while also offering training in entrepreneurship and life skills. Although the program primarily targets women, it also trains a minority of male students (see Appendix Figure A2, panel (b)).

Through these partnerships, we collect administrative data and survey 182 trainees enrolled in automotive mechanics, carpentry, or welding programs. Of these, 33% specialize in automotive mechanics, 19.8% in carpentry, and 21.2% in welding.[7] Trainees are 21.6 years old on average and enrolled in programs lasting 6 to 24 months, with those longer than one year accredited by the Uganda Directorate of Industrial Training (DIT).

Women make up 37% of our trainee sample; this share falls to 18.8% when excluding trainees from the *Girls With Tools* program—our largest provider of female candidates. This nonetheless represents an upward trend compared to the 11% share of female vocational training applicants in welding and motor mechanics documented by Alfonsi et al. (2020). Nationally, in 2023, women were 8.8% of trainees who completed certification in our target sectors. Our sample differs in that it also includes shorter, non-DIT programs and focuses on the Kampala metropolitan area.

2.3 Firm Tasks, Worker Attributes, and Gender Mix

2.3.1 The Multitasking Firm

Firms in this setting are organized around bundles of tasks rather than specialized production activities. Bassi et al. (2023) document that small firms in Uganda exhibit neither

[7] An additional 16.5% are enrolled in broader fields such as electrical work but report job aspirations aligned with our target sectors. About 9.5% of the trainees are enrolled in multiple programs.

horizontal nor vertical specialization. Combined with reliance on probationary workers and apprentices, this lack of specialization requires employers themselves to multitask, combining technical work with supervision and on-the-job training.

Monitoring constraints naturally emerge from the production environment. Employers in our sample spend an average of 10.4 hours at the firm each day, dedicating 2.06 hours to monitoring—the second most time-consuming activity after technical work, and more than they spend on training workers. The physical layout of workshops compounds the challenge: firms typically combine indoor workspaces with large outdoor areas (see Figure A2, panel (a)), making it difficult to observe all workers at once. Over 90% of employers report wanting to increase monitoring, and 83% have experienced theft by an employee.

Piece-rate pay offers a partial solution. In line with previous work in comparable sectors and context Bassi et al. (2023), four out of five employers in our sample report paying piece rates, which tie compensation to output and make effort contractible. However, piece rates address only one margin. They incentivize productivity but do nothing to prevent theft or ensure good behavior when the employer's attention is elsewhere. The core moral hazard problem—can this worker be trusted not to steal when unsupervised?—remains. Anecdotally, employers partly mitigate this risk through network-based hiring, where referrers provide implicit guarantees for those they recommend.

2.3.2 Skills and Gender-Mix Preferences

Consistent with this multi-dimensional production structure, employers do not evaluate workers along a single dimension of "ability" or technical skills. Instead, they value multidimensional workers who combine technical skills (e.g., work quality, learning ability) with non-technical traits related to behavior, intrinsic motivation, and, in particular, trustworthiness.

When asked to identify the most important trait in a worker, employers most frequently cite good behavior—described as trustworthiness, honesty, and not stealing (Figure 1, panel (a)). Good behavior is also the trait employers report as hardest to find and hardest to monitor, as shown in Figure 1, panel (b). Effort and hard work rank second in importance, but only 34% of employers cite effort as hard to monitor—consistent with the prevalence of piece-rate pay, which makes effort contractible. Finally, employers mention learning interest (14.24%) and, in a few cases, cooperation. Interestingly, physical strength is never mentioned in employers' open-ended responses.

Employers hold gendered perceptions of both technical and non-technical skills. When asked to compare men and women on the traits they value most, their responses reveal strong gender stereotypes: men are perceived as stronger in technical domains (e.g., work quality, interest, and learning ability), while women are perceived as superior in trust-related traits—particularly good behavior, described as trustworthiness, honesty, and not stealing (Figure 1, panel (c) and (d). Employers thus appear to perceive women as having a comparative advantage in the traits that are most valued, hardest to find, and hardest to monitor.

We next investigate employers' stated preferences for gender composition in the workplace.[8] When asked, "What is the best gender composition of the workers in your firm?"—on a scale from 0 to 10 men out of 10—84.9% of employers report preferring at least one woman on a ten-person team, and 70.5% prefer two or more (Figure 2, panel (a)). 14.6% state a preference for a roughly gender-balanced team of 4 to 6 women out of 10 workers. Yet at the time of the interview, 86.3% of firms employed an all-male workforce. The gap between stated preferences and actual employment—with 92.3% of

[8]From a policy perspective, this question is key: if employers expressed no interest in female workers, demand-side or matching interventions are unlikely to gain traction.

employers reporting they would prefer more women than they currently have—suggests substantial unmet demand for female workers.

When asked why they do not employ as many women as they would like, employers point to supply-side constraints. Employers estimate that fewer than 5% of the applicant pool are women—on average, 0.5 out of 10. As shown in Figure 2, panel (b), over half cite the scarcity of female applicants as the primary barrier. This is consistent with employers' reliance on personal networks—friends, family, clients, or current employees—for over 80% of their hires, which typically reinforces gender imbalances in male-dominated sectors (Beaman et al., 2018; Alfonsi and De Souza Ferreira, 2024).

Employers also report concerns about hiring women, though these appear secondary to supply constraints. The most frequently cited worry is safety and harassment, mentioned by over 40% of employers (Figure 2, panel (b)).[9] Some employers cite physical strength (24.6%) and effort (20.2%) as concerns, yet, as anticipated earlier, no employer mentions strength as a relevant skill for their business—suggesting these reflect stereotypes rather than job-relevant considerations. Such negative stereotypes appear most relevant for the minority of employers (about 15%) who explicitly state they do not want to hire women. Interestingly, concerns about turnover are not notably stronger for women (mentioned by 12.2% of employers), which we interpret in light of the high baseline turnover rates for workers of both genders.

Employers' stated preferences suggest a substantial unmet demand for female workers. While there is little indication of strong experimenter demand effects—for example, the absence of bunching at focal values such as 1 or 5—these preferences, beliefs, and barriers remain self-reported and do not reveal whether employers would in fact hire women when faced with concrete hiring choices. In these male-dominated sectors, women account for

[9] Question: "If you think of hiring a woman, what do you worry about?".

a much smaller share of the applicant pool, and employers are rarely observed choosing between male and female candidates.[10] We therefore next examine unmet demand for female workers in an incentive-compatible experimental setting.

Employers' stated preferences suggest substantial unmet demand for female workers. While there is little indication of strong experimenter demand effects—for example, the absence of bunching at focal values such as 1 or 5—women account for fewer than 5% of the applicant pool, and employers are therefore rarely observed choosing between male and female candidates. We therefore examine unmet demand for female workers in an incentive-compatible experimental setting.

3 Experimentally Measuring Demand for Female Workers Under Business as Usual

3.1 Experimental Design

We design an experiment to address the female workers supply constraint by providing employers with a pool of vocationally trained candidates that includes women and measuring revealed preferences for hiring male versus female workers.

We use an incentivized resume rating design (Kessler et al., 2019). Employers seeking to hire new workers evaluate hypothetical trainee profiles and, based on their ratings, receive

[10] We lack data on the actual gender composition of these employers' applicant pools. For comparison, Ashraf et al. (2025) finds that firms in India and Turkey draw from applicant pools where men outnumber women four to one. In our sample, employers estimate that 0.5 out of 10 applicants are women.

referrals of real candidates for probationary hiring at their firms. Our primary outcomes are the share of female candidates selected and the gender gap in hiring.[11]

Profiles: Gender and Ability Randomization. To construct the hypothetical trainee profiles, we collect administrative data on 472 enrolled trainees from the partner VTIs in Kampala. Based on this information, we create 36 base profiles by cross-randomizing attributes such as age, marital status, possession of a driver's license, language spoken, motivation for career choice, education, training background, and references. We pilot-test these profiles with 25 employers out of sample.

We then generate four versions of each base profile by cross-randomizing two additional attributes: gender (signaled by an avatar) and technical ability (indicated by class rank and visualized through a star rating—three stars for average performance, five stars for top performance). This yields 144 unique profiles. Each employer is randomized to review one version of each profile, ensuring within-employer variation in gender and ability. Figure 3 illustrates how profiles were presented to employers.

Each employer evaluates 24 randomly selected hypothetical candidate profiles during regular work hours.[12] Our primary outcome is *Meet*: whether the employer wants to initiate a probationary hire ("Do you want us to refer a similar worker to start a probation period at your firm?"; Yes/No). Secondary outcomes probe employers' perceptions of work quality (rated 0–10), behavior and trustworthiness (rated 0–10), and expected monthly earnings one year from now.[13]

[11] During probation, the wage structure typically combines a small daily wage paid by the employer with a training fee paid by the trainee. However, there is substantial heterogeneity across firms, and some firms pay their trainees a fixed wage or piece rate. Thus, we do not focus on gender wage gaps.

[12] We calibrate the number of evaluations to reflect the average monthly inflow of jobseekers (24.6) reported by employers during scoping activities.

[13] *Work Quality*: How would you rate the worker's technical skills and work quality? Please rate on a scale from 0 to 10, where 0 is very low quality and 10 is very high quality."; *Behavior*: How would you rate the workers' behavior (trustworthiness and honesty)? Please, rate on a scale from 0 to 10, where 0

This design allows us to evaluate differential responses by gender and ability in 2×2 design, with random variation across employers for each profile. Each employer sees one randomly selected version for each profile.

Incentives. The IRR design provides incentives for decision-makers to rate candidate profiles in line with their preferences by linking ratings to subsequent referrals while avoiding deception. While employers are told that the profiles are hypothetical, they are also informed that, at the end of the study, referrals will be made from a pool of real candidates drawn from our partner vocational training institutes. This creates a direct link between employers' evaluations and the potential to receive actual trainees. Before evaluating profiles, each employer is guided through this incentive structure by explaining that the more accurately they rate candidates, the better the match between their preferences and the trainees referred to them.

A potential concern is that, given that the mapping from ratings to referrals is not fully disclosed, observed ratings may partly reflect strategic behavior rather than pure preferences (Litwin and Low, forthcoming). In principle, decision-makers could adjust ratings to manage risk—for example, favoring profiles perceived as safer to limit downside outcomes or emphasizing riskier profiles in the hope of receiving an exceptionally good candidate. Such strategic considerations are only problematic if they change employers' ratings differentially by gender. This is unlikely in our setting. Because candidate characteristics are independently and randomly assigned across profiles, conditional on gender, employers do not need to downweight gender in order to obtain referrals of candidates who are more

is not at all trustworthy/honest and 10 is very trustworthy/honest." *Earnings*: What is your best guess of the monthly earnings of this worker a year from now?". As discussed in Appendix A1, we initially preregistered two primary outcomes, the second being an indicator of the desire to make an offer. Due to a programming error, this outcome was not elicited in the first wave (motor mechanics); consequently, we chose not to collect it in the second wave (welding and carpentry) either.

or less risky along other dimensions. We discuss the implications further in the results section.

3.2 Relative Demand for Female Workers

The main sample consists of 906 employers who, as preregistered, pass attention checks related to treatment assignment and exhibit variation in profile evaluations. [14] To investigate the relative demand for female workers, we focus on the 326 employers randomly assigned to the business-as-usual arm (PC). In parallel, we examine heterogeneity by employers' baseline stated preferences for workforce gender mix as a sanity check. We expect demand for female workers to be driven primarily by employers who state any preference for hiring women.

Most employers show a willingness to select female candidates in the experiment: 89.6% of employers select at least one female trainee profile. Among all selected candidates (47% of the total candidates profiles shown), the average share of women is 42.2% (median 44.1%), as shown in panel (a) of Figure 4. The distribution of female candidates selected is highly heterogeneous: while 10.4% of employers select no female trainees, about 35% of employers select more female than male applicants. Throughout, we refer to these selection choices as relative demand for female workers.

Given that approximately 49% of the selected profiles are women, these selection rates imply a gender gap in hiring against women. At the same time, the magnitude of this gap does not appear large relative to the context of male-dominated sectors. We estimate the hiring gender gap, defined as the difference in selection rates between male and female candidates with otherwise identical trainee profiles, using the following regression model:

[14] Sample selection based on the pre-registered protocol is uncorrelated with treatment assignment (Appendix Table A1).

$$\text{Meet}_{ijs} = \beta_0 + \beta_1 \text{Female}_{ij} + \delta_i + \sigma_s + u_{ijs}. \tag{1}$$

Meet_{ijs} denotes employer j's interest in meeting candidate profile i to hire them on probation. Table 3, Panel A, summarizes the results. Female_{ij} equals 1 if employer j is randomly assigned to the female version of profile i. The coefficient β_1 captures the gender gap, measured as the differential effect for female relative to male candidates. δ_i and σ_s denote profile and strata (enumerator, wave, and sector) fixed effects. Standard errors are clustered at the employer and profile levels.

As shown in column 1 of Table 3, Panel A, women are 10.3 percentage points less likely to be selected than men with otherwise identical profiles. This hiring gender gap against women—equivalent to an increase of approximately 1.3 GPA points—is statistically significant (p-value < 0.001) but relatively modest given the extreme gender segregation of the Ugandan labor market and that only 2.3% of workers in these sectors are women.[15] Relative to the mean selection rate, this corresponds to a gap of approximately 22%. For comparison, recent experimental work on discrimination in male-dominated fields in Bangladesh finds gender gaps exceeding 40% (Buchmann et al., 2023), suggesting that the gap is comparatively moderate.

3.3 Heterogeneity by Stated Preferences for Hiring Women

We also estimate the correlation between stated and revealed selection of female workers and re-estimate equation 1 allowing the gender gap to vary by quintiles of stated preferences. Stated preferences for workforce gender mix are positively correlated with the

[15] We estimate the GPA comparison using the random assignment of high and low technical skill signals, indicated by GPA scores of three out of five and five out of five, respectively.

employer's selection of female candidates in the experiment. A 10-percent increase in the share of female workers the employer states to prefer predicts a 4.2 percentage point increase in the likelihood of selecting female workers (p-value < 0.001).

The hiring gender gap narrows significantly with employers' stated preferences for a more gender-balanced workforce, as evident from Figure 4, panel (b). In the regression framework, a 10% increase in the share of female workers in an employer's ideal workforce composition is associated with a 23.8% smaller gender gap relative to employers who report all-male ideal workforce composition (p-value < 0.001).[16] Strikingly, we detect no statistically significant gender gap against female candidates when focusing on employers in the top quintile, those with a stated ideal gender mix of at least 40% women. Thus, in our experiment, about one-fifth of employers exhibit gender-neutral hiring choices under business-as-usual conditions.

3.4 Discussion and Robustness

The relatively high demand for female workers we find in the experiment is unlikely to be an artifact of our design. First, the hiring patterns by gender vary meaningfully with stated preferences for the gender composition of the workforce.[17] As expected, the gender gap is widest among employers who state no preference for hiring women: these employers are five times more likely to select no female candidates in the experiment.[18]

[16]See Appendix Table A9, Column (1).

[17]Similarly, an analysis of secondary outcomes in Appendix Table A5 confirms that employers' stated gendered perceptions of skills and attributes are also reflected in candidate ratings: holding other attributes constant, female profiles are perceived as more trustworthy but less technically able.

[18]Interestingly, 33 employers in PC who currently employ no women and express a preference for a 100% male workforce nonetheless select at least one female profile in the experiment. When asked to explain their potentially inconsistent choices, over 30% cite honesty and trustworthiness as the reason for selecting that female candidate; the remainder state they chose female applicants to give them a chance to prove themselves. We return to this subsample later.

A second concern could be experimenter demands: employers may have overstated preferences for hiring women if they expected these choices to please the research team. Such social considerations may override the referral incentives. We directly elicit employers' beliefs about the research team's preferences regarding their hiring choices, an approach inspired by De Quidt et al. (2018). Most employers (66.7%) state that they believe the team is indifferent, while 23.5% believe we prefer they would hire men; only 6.8% believe we prefer them hiring women. Consistently, awareness of affirmative action policies is limited: only 18.7% of employers had heard of the UNDP's "Gender Equality Seal," the only such policy we could identify in this context.

A final concern relates to the referral incentives themselves, and in particular to employers' expectations about the size and composition of the referral candidate pool. Because women are fewer than men and employers may anticipate competition for female candidates, some employers may refrain from selecting women if they expect referrals of female candidates to be of lower average quality. This concern does not threaten our interpretation: if present, it would mechanically bias measured demand for women downward in the experiment. Consistent with this logic, one would also expect candidate quality to be a strong predictor of selection among female profiles.

Importantly, even if some employers were to select women strategically—for example, to hedge against uncertainty or to increase the likelihood of receiving at least one female candidate—such behavior still reflects a willingness to hire women when provided the opportunity, which is the object of interest. Measured demand for women could only be artificially inflated if employers believed that vocationally trained men were, on average, lower quality and therefore avoided selecting men strategically. This channel is unlikely in our context: vocational training programs are overwhelmingly male-dominated, and vocationally trained candidates—particularly men—are generally perceived as high quality by employers.

We further note that employers' preferences over candidates are highly heterogeneous along multiple dimensions—including gender, age, and technical skills—and that all candidate characteristics are cross-randomized by gender. In addition, referrals are implemented separately by VTI and stratified geographically, which further limits competition for specific candidate types. As a result, it is difficult to conceive of a selection strategy that would require overstating preferences for hiring women in order to obtain otherwise desirable candidate attributes.

In sum, our experimental results reveal substantial unmet demand for female workers in an incentive-compatible setting: when provided the opportunity, most employers select female trainees at meaningful rates. As a result, the gender gap in selection is modest, and a non-trivial share of employers make gender-neutral choices. These findings speak directly to the question raised by Heath et al. (2024) of whether employers in male-dominated sectors are willing to hire women once supply frictions are relaxed.

Next, we investigate to what extent the multidimensional nature of firm tasks and employers' demand for worker attributes shape gender gaps in hiring. Observed selection patterns may reflect either genuinely low bias or the fact that women are relatively advantaged in traits that are particularly valuable under prevailing frictions. However, if women have a comparative advantage in trustworthiness, how can we explain the low hiring rates in equilibrium? We turn to this question in the next section.

4 Frictions and the Hiring Gender Gap

In this section, we examine how relaxing specific workplace frictions affects hiring choices and the gender gap. Our framework emphasizes that firms operate in a multitasking environment in which monitoring and safety frictions shape the relative importance of different worker attributes. Our intuition is that, because non-technical skills and attributes are difficult to observe (Deming, 2017), employers are likely to rely on identity-based beliefs to infer them, and as a result, skills carry identity connotations. The framework therefore implies that reducing monitoring frictions can shift the relative demand for worker types—and, in turn, affect demand for female workers. We next present the experimental design that we leverage to test this implication.

4.1 Treatment Assignment

To test this implication, we randomly assign employers to one of three experimental conditions: a Pure Control arm and two monitoring support regimes. In the Pure-Control (PC) arm, employers make hiring choices under business-as-usual conditions, allowing us to measure demand for female workers and the gender gap in probationary hiring once supply frictions are reduced.

In the two monitoring regimes, firms receive weekly unannounced visits from our team during the probation period. These visits provide external monitoring support, with the goal of reducing employers' monitoring burden. The approach mirrors standard VTI practice, where staff periodically check on trainees during internships to verify attendance. Employers are explicitly informed that assignment to receive visits is random and does not reflect any characteristics of the firm or workers. They are also told that workers will be aware of the monitoring regime, ensuring that any effects operate through expectations of monitoring rather than ex-post detection alone. Importantly, both interventions are framed as gender-neutral by design.

4.1.1 Monitoring Regimes

The two monitoring-support regimes differ in the dimension they target. Employers assigned to the Monitoring-Behavior (MB) arm receive unannounced visits aimed at discouraging trainee misbehavior and preventing theft—the primary monitoring concern reported by employers (Figure 2, panel (a)) and the dimension along which women are perceived to have a comparative advantage. Employers randomly assigned to the Monitoring-Safety (MS) arm receive unannounced visits focused on ensuring that workers are safe (not harassed) and treated with respect. This second intervention targets the main concern employers report when considering hiring women (Figure 2, panel (b)), while again remaining explicitly gender-neutral to avoid priming. [19]

Behavior audits are designed to reduce monitoring frictions related to misbehavior and theft. Employers' expected audit duration—approximately 1.7 hours per audit visit—closely matches their reported desire for extra monitoring, suggesting that the intervention could meaningfully relax monitoring constraints. When monitoring frictions are reduced, the relative value of hard-to-observe trust-related traits declines. Under the assumption

[19]PC script: "We are committed to ensuring that managers and workers will have a positive experience if they end up being matched."; MB script: "We are committed to ensuring that managers and workers will have a positive experience if they end up being matched. Our team members will conduct unannounced weekly visits to some of the firms where new workers are placed during the probation period. These workers will also be informed about the visits. Your firm may be randomly selected to receive such support visits via a lottery. Receiving visits does not reflect any characteristics of the firm or the workers. These visits discourage new workers from practices such as stealing, dishonesty, and disrespect by way of monitoring."; MS script: "We are committed to ensuring that managers and workers will have a positive experience if they end up being matched. Our team members will conduct unannounced weekly visits to some of the firms where new workers are placed during the probation period. These workers will also be informed about the visits. Your firm may be randomly selected to receive such support visits via a lottery. The visits received do not reflect the characteristics of the firm or the workers. Visits are to ensure that the new workers are safe (not harassed) and treated with respect. These visits cannot discourage new workers from practices such as stealing, dishonesty, and disrespect by monitoring."

that women are perceived to have a trustworthiness advantage, we predict that behavior audits will reduce relative demand for female workers, leading to a widening of the gender gap in probationary hiring under MB compared to business-as-usual conditions (PC).

The effect of safety audits on the hiring of women is ex ante more ambiguous and thus particularly policy relevant. Indeed, improving workplace safety—especially for women—is an active policy concern (ILO, 2019), but how employers respond to such interventions in hiring decisions is unclear. Safety audits imply scrutiny of the workplace: coworkers and clients, as well as the employer. If harassment risk comes primarily from coworkers or clients, external oversight addresses employers' safety concerns and may increase employers' willingness to hire women. However, if employers themselves are a source of harassment, safety audits introduce scrutiny of their own behavior—which could discourage them from hiring women. While under both scenarios safety audits reduce harassment, the net effect on demand for female workers is an empirical question.

An additional mechanism consideration is that external scrutiny may be induced, to a lesser extent, also by behavior audits and could therefore drive the potential reduction in the relative demand for female workers. Comparing the effects of Behavior (MB) and Safety (MS) audits provides a benchmark for the presence and magnitude of such scrutiny effects, allowing us to isolate the mechanism operating through reductions in monitoring frictions.

4.1.2 Covariate Balance and First Stage

Table 2 shows covariate balance across the three experimental arms, including employers' sociodemographic characteristics and firm-level attributes.[20] We also assess the overall

[20] We find mild imbalances in one baseline variable: employer VTI training. Results are robust to including this control.

effect of the monitoring regimes on labor demand, measured by the number of profiles selected for probation meetings. Since both audit treatments are designed to ease monitoring, we expect them to increase the rate of selection. Indeed, as shown in Appendix Table A6, employers are 4.1% more likely to select trainees in the Monitoring-Behavior arm (p-value = 0.090) and 6.4% more likely in the Monitoring-Safety arm (p-value = 0.012).

4.1.3 Empirical Strategy

Our main analysis examines the effect of the monitoring regimes on employers' hiring choices, focusing on the gender gap in probationary hiring. To identify the average treatment effect of monitoring regimes on the gender gap, we estimate the following regression model on the main sample of employers:

$$\begin{aligned}\text{Meet}_{ijs} = \beta_0 + \beta_1 \text{Female}_{ij} + \beta_2 \text{MB}_j + \beta_3 \text{MS}_j \\ + \beta_4 \text{Female}_{ij} \cdot \text{MB}_j + \beta_5 \text{Female}_{ij} \cdot \text{MS}_j + \delta_i + \sigma_s + u_{ijs},\end{aligned} \tag{2}$$

which mirrors the baseline specification in equation 1. MBj and MSj are indicators for assignment to the Monitoring-Behavior and Monitoring-Safety arms, respectively; the Pure-Control arm is the omitted category.

We also estimate a fully interacted version of this model to assess heterogeneity in treatment effects by employers' stated preferences for the gender composition of their workforce. This heterogeneity analysis is informative because employers with no stated preference for hiring women are five times more likely to select no female trainees in the experiment, leaving limited scope for monitoring interventions to reduce the gender gap through increased female selection.

4.2 Do Behavior Audits Increase the Gender Gap?

We compare the selection rate of male and female trainees between employers randomly assigned to receive behavior audits (MB), a gender-neutral technology that reduces monitoring costs, and those in the business-as-usual condition (PC).

The raw data, in Appendix Figure A3, show that male trainee selection rates increase under MB relative to the business-as-usual condition, while the average selection rate of female trainees declines. In the regression specification, employers in the Monitoring-Behavior arm are 16.4 percentage points less likely to hire a female trainee, conditional on them being otherwise identical to a male trainee (Table 4, column 1). This treatment effect captured in Figure 5 reflects a statistically significant 63.11% increase in the gender gap compared to the Pure Control group (p-value = 0.008). In terms of the equivalent number of GPA points, the resulting gender gap would equal a 2.1 GPA point increase, compared to 1.3 GPA in the Pure Control arm.

As expected, the effects of behavior audits on the gender gap are larger among employers with stronger stated preferences for hiring women, suggesting that reduced monitoring costs reveal latent barriers that even pro-diversity employers face when evaluating female candidates. In the subsample of employers with a stated demand for hiring female workers, the gender gap widens significantly more (+113.2%), going from 7.8 percentage points in the Pure Control to 15.5 percentage points under Monitoring Behavior.

Strikingly, the effect of behavior audits on the gender gap is strongest among employers who state a preference for a nearly gender-balanced workforce. For this group, the gender gap is not statistically different from zero in the Pure Control arm but increases to 15.2 percentage points under behavior audits (p-value < 0.001). This treatment effect is 2.7 times larger than the baseline estimate for the full sample. These employers—who appear gender-neutral under business-as-usual conditions—exhibit gender gaps comparable to those employers with no stated preferences for hiring women once monitoring frictions are reduced.

To summarize, support in monitoring workers' behavior and preventing theft reduces the relative demand for women and widens the hiring gender gap, revealing latent gender gaps, especially among employers who appeared gender-neutral based on their stated preferences for hiring women. A simple back-of-the-envelope calculation indicates that the demand for trustworthiness explains approximately 60% of the correlation between stated diversity preferences and the gender gap under business-as-usual conditions (Appendix Table A9).

4.3 Do Safety Audits Affect the Gender Gap?

We next examine how safety audits affect the gender gap in hiring, comparing employers in the Monitoring-Safety arm to those operating under business-as-usual conditions (PC). As discussed earlier, the predicted effect of safety audits on the gender gap is theoretically ambiguous. On the one hand, increased external oversight could raise perceived scrutiny and compliance costs, potentially reducing employers' willingness to hire women. On the other hand, safety audits may reassure employers by reducing concerns about workplace safety and harassment, thereby reducing the relative cost of selecting female candidates.

The data clearly support the latter mechanism. As shown in the raw data (Figure A3), safety audits increase the relative selection of female trainees, implying a reduction in the gender gap against women. Consistent with this pattern, regression estimates of β_5 in Table 4, column (1), indicate that assignment to the Monitoring-Safety arm reduces the gender gap by 8.6 percentage points on average. We can reject the null hypothesis that safety audits widen the gender gap at the 1% level (p-value = 0.005).

Unlike behavior audits, the effect of safety audits is not concentrated among employers with particular stated preferences for a gender-mixed workforce. Instead, the reduction in the gender gap is broadly shared across the preference distribution, with one exception: among employers who report no preference for hiring women, safety audits have no detectable effect on the gender gap (Table 4, column (3)).

Overall, the magnitude of the effect of safety audits on the hiring gender gap is substantial. Safety support reduces the gender gap by 83.5% relative to business-as-usual conditions (p-value = 0.008), and we cannot reject the null hypothesis that the gender gap under safety audits is statistically indistinguishable from zero. These results indicate that addressing frictions related to workplace safety and harassment —together with a reduction in search costs—closes observable gender gaps in employers' hiring choices in our setting.

4.4 Mechanism Discussion

Taken together, the results are consistent with our predictions: technologies that address workplace frictions change the relative value of worker attributes; because these attributes carry gendered connotations, they directly affect discrimination in hiring. When monitoring frictions are reduced through behavior audits, the relative importance of trustworthiness declines, leading to an increase of the hiring gender gap against women. Conversely, when harassment-related concerns, which are more binding for women in a gender-traditional environment, are addressed through safety audits, observable gender gaps are substantially reduced.

A potential challenge to this interpretation is that monitoring audits may affect hiring indirectly through increased scrutiny rather than through changes in the valuation of worker attributes. For instance, employers in the Monitoring-Behavior (MB) arm might reduce female hiring not because trust becomes less valuable, but because increased monitoring increases the perceived cost of employing women by limiting opportunities for harassment. The Monitoring-Safety (MS) arm provides a natural benchmark to assess this concern, as it introduces direct employer scrutiny. The evidence suggests that such scrutiny effects are unlikely to be the primary driver of the MB-PC gender gap differences, because direct scrutiny under safety audits increases the relative demand for women on average. This pattern contrasts sharply with the effect of behavior audits and indicates that scrutiny alone cannot explain the reduction in demand for female workers under MB.

More broadly, the results are difficult to reconcile with alternative explanations such as customer discrimination (employers actually expect female workers to attract customers), experimenter demand (all technologies are framed as gender-neutral), or strategic behavior induced by referral incentives (referral incentives are the same across arms), reinforcing the interpretation that workplace frictions and the gendered valuation of skills drive the observed patterns. Finally, we examine whether employers perceive women as more trustworthy because they believe women have fewer outside options. However, employers holding this belief are less likely to prefer a gender-mixed workforce (Figure A6).

In summary, under business-as-usual conditions, monitoring frictions make employers place high value on trust-related traits. Because women are perceived as more trustworthy, these frictions raise the relative demand for female workers and mask underlying bias. When monitoring frictions are reduced, this hidden discrimination becomes visible. Why are women not hired in the first place if they have such a comparative advantage? Other frictions—most notably search costs and safety/harassment-related concerns—depress employers' willingness to hire women, helping explain why women remain underrepresented despite apparent demand under certain conditions.

5 Link to Standard Discrimination Models

Our results show that employers attempt to select workers based on perceived non-technical attributes such as trustworthiness, using gender as a proxy, which generates positive statistical discrimination in favor of women along this dimension. This advantage is offset by either prejudice or by beliefs about other productive traits—such as technical ability—so that overall hiring disparities by gender persist. When productivity is multidimensional, however, positive statistical discrimination along one dimension—such as trustworthiness—can mask negative bias along others.

To fully characterize gender disparities in hiring, it is therefore necessary to identify the sources of the negative discrimination faced by women. Under a benign interpretation, the observed gender gap could be efficient if employers are screening for both technical ability and trustworthiness and use gender as a proxy for these traits. However, statistical inference may be based on inaccurate beliefs; in that case, what appears as statistical discrimination may in fact reflect prejudice as emphasized by Lang and Spitzer (2020) and Bohren et al. (2023).

In this section, we first test whether the discrimination we uncover is consistent with statistical discrimination based on ability, leveraging the cross-randomization of candidate gender and technical ability in the worker profiles. We find that the hiring gap against women in our setting is inconsistent with negative statistical discrimination on technical ability. Second, we elicit employers' beliefs on both trainee skills and trustworthiness by gender, and compare them actual differences in trainee behavior. Against employers' beliefs, we find no evidence of gender differences in trainees skills or pro-sociality in our sample, accr

5.1 Testing for Statistical Discrimination Based on Technical Skills Against Female Workers

We implement the standard test of statistical discrimination based on technical skills, leveraging the cross-randomization of candidate gender and technical ability in the worker profiles. If employers use gender as a proxy for ability under imperfect information, gender gaps in hiring should narrow in response to high-ability signals (Bertrand and Duflo, 2017).

We estimate the following model to investigate the role of worker ability in shaping the hiring gender gap within each treatment arm, and across arms:

$$\begin{aligned}\text{Meet}_{ijs} = \beta_0 &+ \beta_1\text{Female}_{ij} + \beta_2\text{HighAbility}_{ij} + \beta_3\text{Female}_{ij}\cdot\text{HighAbility}_{ij} \\ &+ \beta_4\text{MB}_j + \beta_5\text{MS}_j + \beta_6\text{Female}_{ij}\cdot\text{MB}_j + \beta_7\text{Female}_{ij}\cdot\text{MS}_j \\ &+ \beta_8\text{HighAbility}_{ij}\cdot\text{MB}_j + \beta_9\text{HighAbility}_{ij}\cdot\text{MS}_j \\ &+ \beta_{10}\text{Female}_{ij}\cdot\text{HighAbility}_{ij}\cdot\text{MB}_j + \beta_{11}\text{Female}_{ij}\cdot\text{HighAbility}_{ij}\cdot\text{MS}_j \\ &+ \delta_i + \sigma_s + u_{ijs}\end{aligned} \tag{3}$$

HighAbility_{ij} indicates whether profile i presented to employer j includes high-skill signals. We cluster standard errors at the employer and profile level. Our primary ability measure is GPA ranking within the class, based on both theoretical and practical evaluations, shown on profiles as five stars for top performers and three stars for average performers within their training cohort. We also construct a profile quality index that equally weights secondary education completion, DIT certification, enrollment in a 12-month program, and a high GPA.

First, we focus on the business-as-usual condition, which corresponds to how discrimination is typically measured in the standard framework. Although employers place substantial weight on technical skill signals—high-skill profiles are 11.8 percentage points more likely to be selected (25.4%; p-value < 0.001)—female candidates do not benefit differentially from these signals. The interaction between ability and gender is negative and statistically insignificant (p-value $= 0.307$). Thus, the gender gap do not narrow with higher observed ability.[21]

The discrimination against women we uncover in the Monitor Behavior arm is inconsistent with statistical discrimination based on technical ability. If employers were reducing female hiring because they believe women are less skilled, we would expect the effect to be concentrated among lower-ability candidates. Instead, when monitoring constraints are relaxed, employers reduce the hiring of both high-ability and low-ability female candidates by similar amounts. β_{10} in the equation above captures whether changes in the gender gap induced by reductions in monitoring costs—relative to business-as-usual conditions—are different by ability signals. As shown in Table A4, we find no evidence that technical ability accounts for the widening gender gap (p-value $= 0.871$).

Similarly, when we provide safety support, the hiring gap against women does not narrow systematically more for high-skill candidates. Table A2, Panel B, shows that high-skill profiles are 8.1 percentage points more likely to be selected (p-value < 0.001), while the interaction between skill and gender is negative and statistically insignificant (p-value $= 0.343$). This rules out statistical discrimination based on technical ability as an explanation for the gender gap in MS.

[21] For ease of interpretation, we present results based on sample-split estimation. For transparency, we also report estimates from the fully saturated model in Table A4, which yield consistent results. Results are similar using an alternative profile quality index: a one-standard-deviation increase in profile quality raises the probability of selection by 13.2 percentage points (p-value < 0.001) but does not reduce the gender gap (p-value $= 0.454$); see Table A3.

5.2 Beliefs Accuracy

We assess belief accuracy by comparing employers' gendered beliefs about skills and attributes to survey data from 182 vocational trainees across seven training centers, incentivizing employer responses with the true survey results.

Technical skills. We assess two measures of technical skills. To measure practical skills, we ask trainees to self-report their ability to perform key technical tasks (e.g., for trainees in motor mechanics: "conducting an oil change"), following Alfonsi et al. (2020). We assess theoretical knowledge with a set of theoretical questions specific to their specialization. The theoretical questions were designed by VTI instructors; an exam example is in Figure A4.[22] The results of both trainee outcomes and employers beliefs by gender are summarized in Figure 7 and in Appendix Table A11.

Employers perceive female workers as less skilled, both on practical and theoretical tasks. On average, employers expect women to be 20.8% and 25.4% less likely than men to answer correctly on theory and practical exams, respectively (both p-value < 0.001). These beliefs are inaccurate: we find no evidence of gender differences in trainee ability or skills. In self-assessments, female and male trainees are equally likely to report being able to perform a standard sector task (p-value $= 0.831$). On the theory questions, women are, if anything, more likely to answer correctly, though this difference is not significant (10.5 percentage points; p-value $= 0.249$).

Trustworthiness. The literature typically measures trust and trustworthiness using the investment game (Berg et al., 1995), which captures beliefs about how much another

[22] Because of time constraints, we elicited employers' beliefs on one randomly selected survey question and matched these beliefs to actual trainee outcomes for that question.

party will reciprocate trust, conditional on having been trusted. Alternatively, survey-based measures—such as those in the World Values Survey—elicit generalized trust or prosociality. Closest to our setting, Caria and Falco (2024) design an incentive compatible game to measure of moral hazard defined as unsupervised effort in productive tasks. Neither definition is close to the notion of trustworthiness as honesty and non-stealing that we refer to in this paper. Thus, we design a novel incentivized game to measure both effort and cheating, supervised and unsupervised.

The "Cheating" task is as follows. We randomly assign trainees to supervised or unsupervised conditions. In the unsupervised condition, trainees receive a flat wage to complete up to 12 proofreading tasks, with no output verification. In the supervised condition, trainees are paid piece rate, with supervisors verifying output and paying only for completed tasks. The number of completed tasks is our measure of output/effort. We introduce an incentive to cheat: trainees can increase their pay by 50% by claiming to find more than 10 proofreading mistakes, although by design there are fewer than 10, making this a direct measure of cheating. The sample is stratified by treatment and gender to estimate mean outcomes separately for each group.[23] Employers predict average task completion and likelihood of cheating for male and female trainees separately, under both supervised and unsupervised conditions. Results are provided in columns (7)-(11) of Appendix Table A11.

As a sanity check, we note that employers expect supervision to increase task completion by 21.2 percentage points (p-value < 0.001) from a baseline rate of 66.9% (no supervision) on average. However, this effect is driven by male trainee increasing effort under supervision. Indeed, they expect male trainees to complete fewer tasks than female trainees

[23]To avoid deception, individual output is not verified in the unsupervised arm. Effort can only be assessed at the treatment-by-gender cell level.

when unsupervised (4.7 percentage points, p-value < 0.001), however the gap disappears under supervision (p-value < 0.001).

Most relevant to our question, employers expect unsupervised female trainees to misreport less than unsupervised men (11.5 percentage points relative to a perceived average likelihood of 46.6% for men, p-value < 0.001).[24] The effect is driven by both men and women: 25.6% of men cheat when supervised, and women are 5.7 percentage points less likely than men to cheat. Thus, overall employers expect women to exert more effort and cheat less than men, especially when unsupervised.

Employers' expectations are misaligned with actual trainee behavior. Task completion is nearly universal (97%) across gender and supervision. Misreporting is low on average (10.4%), with no statistically significant gender differences; if anything, women are more likely to cheat absent supervision (9.6 percentage points, $p = 0.257$). Supervision substantially reduces misreporting (63%, $p = 0.009$), with no differential effect by gender ($p = 0.415$).

5.3 Measuring Bias as a Residual in a Multi-Dimensional Skill Framework

What is the extent of bias against women in this setting? As noted above, in the standard unidimensional worker skills framework, the gender gap observed under business-as-usual conditions would be interpreted as evidence of a hiring penalty against women of approximately 10 percentage points attributable to bias. We show, however, that this interpretation is incomplete. When productivity is multidimensional, outcome gaps no longer map cleanly into prejudice. Positive statistical discrimination along one productive

[24]Employers overall expect supervision to reduce the overall likelihood of cheating by 18.1 percentage points.

dimension can mask negative discrimination or prejudice along another, while negative statistical discrimination along unobserved dimensions can mechanically inflate residual measures of bias. As a result, residual approaches will offer different estimates of bias depeding on which relevant productivity characteristics are observed.

Our experimental results show that this simple identification problem has first-order measurement implications. In a framework with both technical ability and trustworthiness as productive skills, the implied degree of bias against women is at least 60% larger (16 percentage points) than what would be inferred under a one-dimensional model.[25] In a framework with both technical skills, trustworthiness, and safety concenrs, where we assume that employers' safety concerns are to be seen as a determinant of productivity, however, the estimates of bias goes back down to 8.1 percentage points.[26]

However, what should be considered as a determinant of productivity is not clear ex-ante and conceptually important for the estimate of bias. If employers treat safety risk as a productivity cost, then part of the observed gender gap reflects productive safety concerns, and the implied degree of bias is 8.1 percentage points. If instead employers trade off productivity against concern for women's welfare, as in Macchi and Stalder (2023), the relevant distinction is between altruism and paternalism. Given that all candidates have self-selected into vocational training for male-dominated trades and that female and male trainees exhibit similar risk preferences, employers' responses to reductions in safety-related monitoring frictions are most consistent with paternalistic discrimination (Buchmann et al., 2023). Under this interpretation, the full 16 percentage points gap

[25]Because monitoring frictions are not fully eliminated, this estimate should be interpreted as a lower bound.

[26]Our experiment is useful because it provides variation on a third dimension: safety concerns. Comparing the Monitoring-Safety (MS) and Pure-Control (PC) arms allows us to estimate the share of the business-as-usual gender gap attributable to safety-related considerations. Assuming safety concerns are similar under PC and Monitoring-Behavior (MB), approximately half of the gender gap revealed when monitoring frictions are reduced can be attributed to safety concerns (8.65 percentage points). This estimate should be interpreted as an upper bound, as any external audit is likely to reduce safety concerns relative to business as usual.

revealed when monitoring frictions are reduced can be viewed as bias deriving in part from gender norms that promote the protection of women.

More broadly, this exercise highlights that the technology under which hiring decisions are made—and, critically, our assumptions about which attributes constitute productivity—have drastic implications for how discrimination is measured. In our setting, depending on whether trustworthiness and safety-related considerations are treated as productive, the estimated degree of residual bias against women ranges from as little as 1.9 percentage points to as much as 16.75 percentage points. Of course, there may be additional dimensions of productivity that we do not observe. As a result, outcome gaps are not necessary for the presence of discrimination: positive statistical discrimination along some dimensions can mask negative discrimination along others. More generally, when productivity is multidimensional, estimates of bias are inherently sensitive to assumptions about which worker attributes are productive.

6 Conclusions

We study hiring preferences by gender and show evidence of an unmet demand for female workers in three male-dominated sectors in Uganda. We find that labor market frictions shape the demand for skills and, in turn, the extent of gender discrimination in hiring.

In a multi-dimensional framework of worker skills (Deming and Silliman, 2025; Woessmann, 2024), our results indicate that the absence of a gender gap in hiring does not imply the absence of discrimination. Positive selection along some traits can offset negative bias along others, masking underlying discrimination. In our setting, women's perceived advantage in trustworthiness compensates for safety concerns as well as for potential bias on other dimensions. This mechanism is conceptually related to, but distinct from, Ashworth et al. (2024); Bohren et al. (2019), showing that discrimination can induce stronger average skills among the selected members of disadvantaged groups; rather than operating through biased selection over time, it arises from the interaction of multiple skill dimensions and the frictions that determine their relative importance.

One interpretation of our results close to Becker (1957) is that, under some frictions, discrimination behaves like a "luxury" good — whereby, employers face more constraints and discriminate less. However, our findings also suggest that the distinction between taste-based a la Becker and statistical discrimination may be less central for interpreting outcome gaps. The fact that residual discrimination may partly reflect statistical discrimination along unobserved, potentially infinite, productive dimensions does not imply that such patterns are efficient or benign. Beliefs can be inaccurate (Bohren et al., 2023) and employers' beliefs are themselves shaped by norms (Buchmann et al., 2023) or past discrimination (Bohren et al., 2019, 2025).

Implications. Our results highlight how firms' production technologies shape which attributes are valued. These technologies are not exogenously given and can carry important identity implications. An important implication is that technologies—and more broadly, labor market interventions aimed at reducing frictions in low-income settings—may have unintended distributional consequences, including effects on gender inequality.

In line with Bandiera et al. (2022), our results also indicate that market conditions shape the success of interventions by influencing employer attitudes toward women. These interactions complicate predicting the effectiveness of training interventions across different settings. Structural barriers—such as search frictions and safety concerns—must be addressed to sustainably expand opportunities. More constructively, our evidence—consistent with Field and Vyborny (2022) and Buchmann et al. (2023)—suggests that relatively simple safety interventions can meaningfully reduce the perceived safety cost of hiring women. Finally, Shifting hiring practices from informal, referral-based systems common in smaller firms to more structured recruitment processes prevalent in larger firms may also be an effective tool to broaden candidate pools and reduce occupational segregation.

External validity. Our results are most directly relevant to frictional labor markets in low-income settings, particularly sub-Saharan Africa. However, the interaction between skills and gender stereotypes is not unique to our context.

Evidence from the United States shows that beliefs about prosociality are often systematically distorted along gender lines (Exley et al., 2025). Moreover, macro-estimates show that technical change reshaped the demand for skills with substantial implications for gender and racial wage gaps in the past decades (Caunedo and Keller, 2022).

History illustrates similar dynamics. During World War II, labor scarcity brought many women into analytical fields such as cryptanalysis, where perceived strengths in precision and patience were valued.[27] Yet after the war, women's labor force participation reverted to prewar levels. While causal identification is difficult, the literature attributes most of this reversion to declining demand rather than changes in female labor supply (Goldin, 1991; Rose, 2018)—suggesting demand shifts were related to changes in production needs that temporarily raised the demand for women's skills rather than a lasting shift in preferences for diversity.

[27] "It was generally believed that women were good at tedious work—and, as code-breaker Ann Caracristi recalled, the early stages of cryptanalysis were very tedious" (Hammond, 2017). The stereotype that casts men as geniuses and women as hardworking persists in today's academia (Eberhardt et al., 2023).

References

Agte, Patrick and Arielle Bernhardt, "The economics of caste norms: Purity, status, and women's work in India," *mimeo*, 2024.

Alfonsi, Livia and Pedro De Souza Ferreira, "Whom would you rather work with? Uncovering the Referral Gender Gap," 2024.

_ , Mary Namubiru, and Sara Spaziani, "Gender gaps: back and here to stay? Evidence from skilled Ugandan workers during COVID-19," *Review of Economics of the Household*, 2024, *22* (3), 999–1046.

_ , Oriana Bandiera, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali, "Tackling youth unemployment: Evidence from a labor market experiment in Uganda," *Econometrica*, 2020, *88* (6), 2369–2414.

Arrow, Kenneth J., "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton: Princeton University Press, 1973, pp. 3–33.

Ashraf, Nava, Oriana Bandiera, Virginia Minni, and Victor Quintas-Martínez, "Gender Gaps across the Spectrum of Development: Local Talent and Firm Productivity," 2025. Revise and Resubmit at *Review of Economics and Statistics*.

Ashworth, Scott, Christopher R Berry, and Ethan Bueno de Mesquita, "Modeling theories of women's underrepresentation in elections," *American Journal of Political Science*, 2024, *68* (1), 289–303.

Bandiera, Oriana, Ahmed Elsayed, Andrea Smurra, and Céline Zipfel, "Young adults and labor markets in Africa," *Journal of Economic Perspectives*, 2022, *36* (1), 81–100.

Bassi, Vittorio, Jung Hyuk Lee, Alessandra Peter, Tommaso Porzio, Ritwika Sen, and Esau Tugume, "Self-employment within the firm," Technical Report, National Bureau of Economic Research 2023.

Beaman, Lori and Jeremy Magruder, "Who gets the job referral? Evidence from a social networks experiment," *American Economic Review*, 2012, *102* (7), 3574–3593.

__ , **Niall Keleher, and Jeremy Magruder**, "Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi," *Journal of labor economics*, 2018, *36* (1), 121–157.

Becker, Gary S, *The economics of discrimination*, University of Chicago Press, 1957.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, reciprocity, and social history," *Games and economic behavior*, 1995, *10* (1), 122–142.

Bernhardt, Arielle, Erica Field, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer-Moore, "Male social status and women's work," in "AEA Papers and Proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 363–367.

Bertrand, Marianne and Esther Duflo, "Field experiments on discrimination," *Handbook of economic field experiments*, 2017, *1*, 309–393.

__ **and Sendhil Mullainathan**, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American economic review*, 2004, *94* (4), 991–1013.

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg, "The dynamics of discrimination: Theory and evidence," *American economic review*, 2019, *109* (10), 3395–3436.

__ , **Kareem Haggag, Alex Imas, and Devin G Pope**, "Inaccurate statistical discrimination: An identification problem," *Review of Economics and Statistics*, 2023, pp. 1–45.

__ , **Peter Hull, and Alex Imas**, "Systemic discrimination: Theory and measurement," *The Quarterly Journal of Economics*, 2025, *140* (3), 1743–1799.

Boudreau, Laura E, Sylvain Chassang, Ada Gonzalez-Torres, Rachel Heath, and National Bureau of Economic Research, "Monitoring harassment in organizations," Technical Report, National Bureau of Economic Research Cambridge, MA 2023.

Breza, Emily and Supreet Kaur, "Labor Markets in Developing Countries," *Annual Review of Economics*, 2025, *17.*

Brown, Christina, "Understanding Discrimination by Managers," 2023. Unpublished manuscript.

Buchmann, Nina, Carl Meyer, and Colin D Sullivan, "Paternalistic discrimination," Technical Report, Working Paper 2023.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived social norms: Women working outside the home in Saudi Arabia," *American economic review*, 2020, *110* (10), 2997–3029.

Caria, Stefano A and Paolo Falco, "Skeptical employers: Experimental evidence on biased beliefs constraining firm growth," *Review of Economics and Statistics*, 2024, *106* (5), 1352–1368.

Caunedo, Julieta and Elisa Keller, "Technical change and the demand for talent," *Journal of Monetary Economics*, 2022, *129*, 65–88.

Deming, David and Mikko Silliman, "Skills and human capital in the labor market," in "Handbook of Labor Economics," Vol. 6, Elsevier, 2025, pp. 115–157.

Deming, David J, "The growing importance of social skills in the labor market," *The quarterly journal of economics*, 2017, *132* (4), 1593–1640.

Eberhardt, Markus, Giovanni Facchini, and Valeria Rueda, "Gender differences in reference letters: Evidence from the economics job market," *The Economic Journal*, 2023, *133* (655), 2676–2708.

Exley, Christine L, Oliver P Hauser, Molly Moore, and John-Henry Pezzuto, "Believed gender differences in social preferences," *The Quarterly Journal of Economics*, 2025, *140* (1), 403–458.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global evidence on economic preferences," *The quarterly journal of economics*, 2018, *133* (4), 1645–1692.

Fernández, Raquel and Alessandra Fogli, "Culture: An empirical investigation of beliefs, work, and fertility," *American economic journal: Macroeconomics*, 2009, *1* (1), 146–177.

Field, Erica and Kate Vyborny, "Women's mobility and labor supply: experimental evidence from Pakistan," *Asian Development Bank Economics Working Paper Series*, 2022, (655).

Foster, Andrew D and Mark R Rosenzweig, "A test for moral hazard in the labor market: Contractual arrangements, effort, and health," *The Review of Economics and Statistics*, 1994, pp. 213–227.

Gentile, Elisabetta, Nikita Kohli, Nivedhitha Subramanian, Zunia Saif Tirmazee, and Kate Vyborny, "Barriers to entry: Decomposing the gender gap in job search in urban Pakistan," Technical Report, ADB Economics Working Paper Series 2023.

Goldin, Claudia and Cecilia Rouse, "Orchestrating impartiality: The impact of "blind" auditions on female musicians," *American economic review*, 2000, *90* (4), 715–741.

Goldin, Claudia D, "The role of World War II in the rise of women's employment," *The American Economic Review*, 1991, pp. 741–756.

Goldstein, Markus, Paula Gonzalez Martinez, and Sreelakshmi Papineni, "Tackling the global profitarchy: Gender and the choice of business sector," *World Bank Policy Research Working Paper*, 2019, (8865).

Gradín, Carlos, "Occupational gender segregation in post-apartheid South Africa," *Feminist Economics*, 2021, *27* (3), 102–133.

Hammond, Sarah, "The Secret History of the Female Code Breakers Who Helped Defeat the Nazis," *Politico Magazine*, October 10 2017.

Heath, Rachel, "Why do firms hire using referrals? Evidence from Bangladeshi garment factories," *Journal of Political Economy*, 2018, *126* (4), 1691–1746.

_, Arielle Bernhardt, Girija Borker, Anne Fitzpatrick, Anthony Keats, Madeline McKelway, Andreas Menzel, Teresa Molina, and Garima Sharma, "Female labour force participation," *VoxDevLit*, 2024, *11* (1), 1–43.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

International Labour Organization, "ILO Modelled Estimates and Projections database (ILOEST)," https://ilostat.ilo.org/data/ 2025. Accessed January 7, 2025.

Jayachandran, Seema, "The roots of gender inequality in developing countries," *Annual review of economics*, 2015, *7* (1), 63–88.

_, "Social norms as a barrier to women's employment in developing countries," *IMF Economic Review*, 2021, *69* (3), 576–595.

Kessler, Judd B, Corinne Low, and Colin D Sullivan, "Incentivized resume rating: Eliciting employer preferences without deception," *American Economic Review*, 2019, *109* (11), 3713–3744.

Kline, Patrick, Evan K Rose, and Christopher R Walters, "Systemic discrimination among large US employers," *The Quarterly Journal of Economics*, 2022, *137* (4), 1963–2036.

Lang, Kevin and Ariella Kahn-Lang Spitzer, "Race discrimination: An economic perspective," *Journal of Economic Perspectives*, 2020, *34* (2), 68–89.

Litwin, Ashley and Corinne Low, "Measuring Discrimination with Experiments," in Alex Rees-Jones, ed., *Handbook of Experimental Methods in the Social Sciences*, Cheltenham, UK: Edward Elgar Publishing, forthcoming.

Macchi, Elisa and Jeremia Stalder, "Work over just cash: Informal redistribution," *Revise and Resubmit at Econometrica*, 2023.

Macchiavello, Rocco, Andreas Menzel, Atonu Rabbani, and Christopher Woodruff, "Challenges of change: An experiment promoting women to managerial roles in the bangladeshi garment sector," Technical Report, National Bureau of Economic Research 2020.

Organization, International Labour, "Violence and Harassment Convention, 2019 (No. 190)," https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO:12100:P12100_INSTRUMENT_ID:3999810 2019. Adopted at the 108th International Labour Conference, Geneva, 21 June 2019.

Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *American Economic Review*, 1972, *62* (4), 659–661.

Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth, "Measuring and bounding experimenter demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

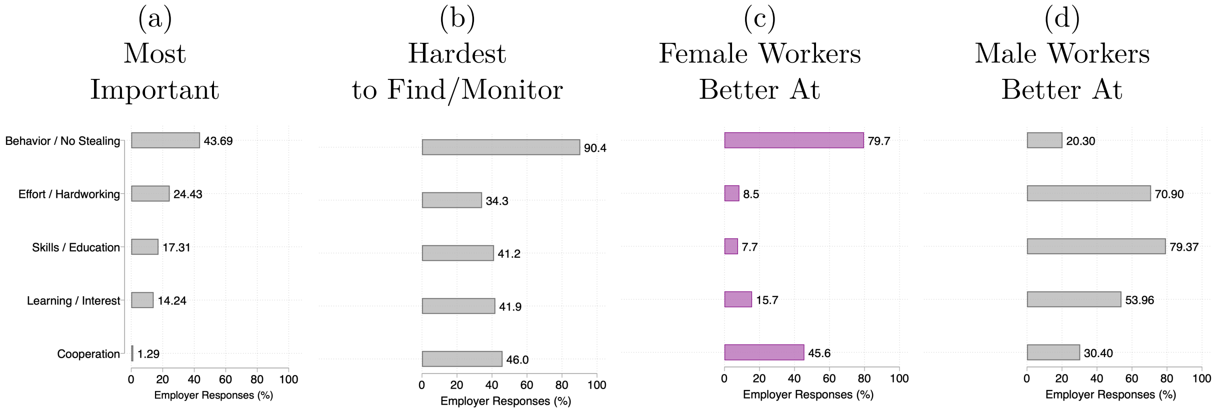
Rose, Evan K, "The rise and fall of female labor force participation during World War II in the United States," *The Journal of Economic History*, 2018, *78* (3), 673–711.

Sharma, Karmini, "Tackling Sexual Harassment: Short and Long-Run Experimental Evidence from India," Technical Report, Mimeo 2023.

Woessmann, Ludger, "Skills and earnings: A multidimensional perspective on human capital," *Annual Review of Economics*, 2024, *17.*

Figures

Figure 1: Multi-dimensional and gendered worker attributes.

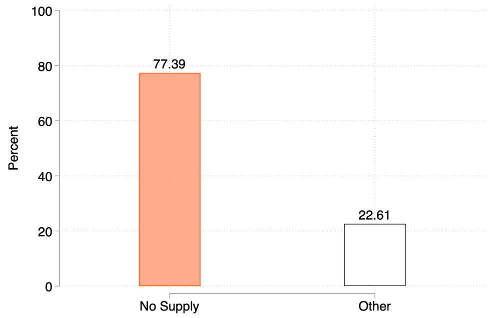


Notes: The figure presents descriptive statistics from the full sample of employers. Panel (a) reports the distribution of employers' responses to the open-ended question, "What is the single most important trait in a worker?", grouped into the standardized categories shown. A residual "Other" category (13.43%) includes answers related to personal characteristics such as age or background. Panel (b) shows the share of employers stating that a given trait is the hardest to find or monitor, for traits derived from Panel (a). Panels (c) and (d) report the share of respondents answering "Female" or "Male" to the question, "Do you think male or female workers are better at [trait]? (Female, Male, or Indifferent)," for traits derived from Panel (a).

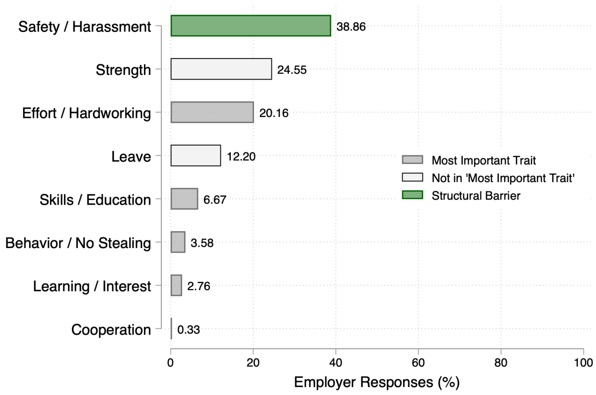
Figure 2: Workforce Gender Mix Preferences and Barriers to Hiring Women

(a) Employers' Desired and Actual Gender Mix Among Workforce

(b) "Can you explain why, in your opinion, you hire no female workers?"

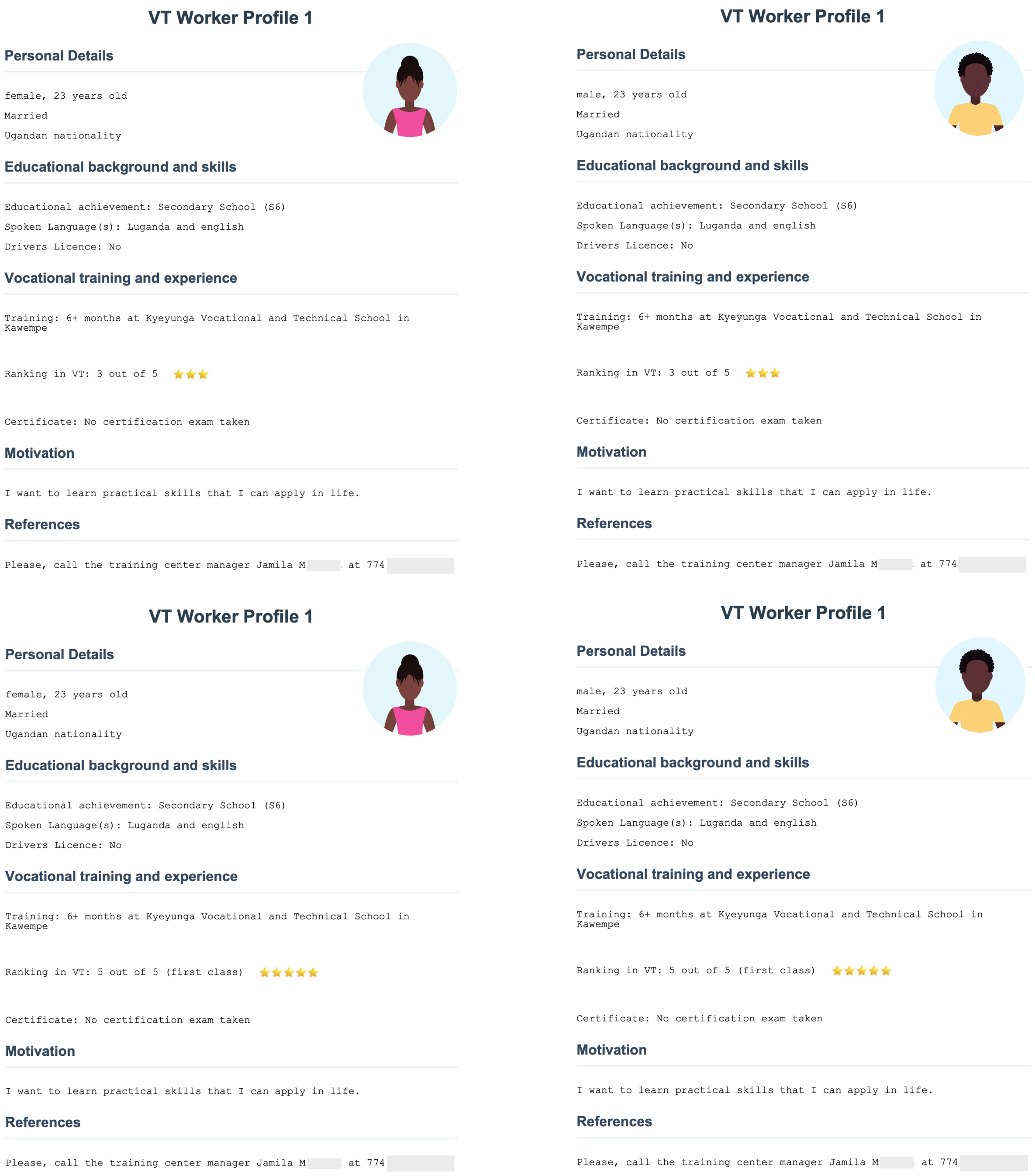


(c) "If you think of hiring a woman, what do you worry about?"



Notes: Panel (a) is based on the employers full sample. Panel (a) presents the distributions of the desired gender composition — on a scale from 0 men out of 10 (all women) to 10 men out of 10 (all men) — versus the actual share of women among current workers, rescaled to 10. Panel (b) and (c) are based on employers who currently do not employ any female workers. Panel (b) plots answers to the question: " Can you explain why, in your opinion, you hire no female workers?". "Other" includes responses related to the following categories, listed by decreasing percentage: "Leave", "Strength", "Effort / Hardworking", "Skills / Education", "Safety / Harassment", "Behavior / No Stealing". This includes only employers who currently do not employ any female workers. Panel (c) shows the distribution of answers to the open-ended question:"If you think of hiring a woman, what do you worry about?". Of these, 10.73% responded "Nothing" and 1.79% "Other".

Figure 3: CV Examples



VT Worker Profile 1

Personal Details

female, 23 years old
Married
Ugandan nationality

Educational background and skills

Educational achievement: Secondary School (S6)
Spoken Language(s): Luganda and english
Drivers Licence: No

Vocational training and experience

Training: 6+ months at Kyeyunga Vocational and Technical School in Kawempe

Ranking in VT: 3 out of 5 ★★★

Certificate: No certification exam taken

Motivation

I want to learn practical skills that I can apply in life.

References

Please, call the training center manager Jamila M at 774

VT Worker Profile 1

Personal Details

male, 23 years old
Married
Ugandan nationality

Educational background and skills

Educational achievement: Secondary School (S6)
Spoken Language(s): Luganda and english
Drivers Licence: No

Vocational training and experience

Training: 6+ months at Kyeyunga Vocational and Technical School in Kawempe

Ranking in VT: 3 out of 5 ★★★

Certificate: No certification exam taken

Motivation

I want to learn practical skills that I can apply in life.

References

Please, call the training center manager Jamila M at 774

VT Worker Profile 1

Personal Details

female, 23 years old
Married
Ugandan nationality

Educational background and skills

Educational achievement: Secondary School (S6)
Spoken Language(s): Luganda and english
Drivers Licence: No

Vocational training and experience

Training: 6+ months at Kyeyunga Vocational and Technical School in Kawempe

Ranking in VT: 5 out of 5 (first class) ★★★★★

Certificate: No certification exam taken

Motivation

I want to learn practical skills that I can apply in life.

References

Please, call the training center manager Jamila M at 774

VT Worker Profile 1

Personal Details

male, 23 years old
Married
Ugandan nationality

Educational background and skills

Educational achievement: Secondary School (S6)
Spoken Language(s): Luganda and english
Drivers Licence: No

Vocational training and experience

Training: 6+ months at Kyeyunga Vocational and Technical School in Kawempe

Ranking in VT: 5 out of 5 (first class) ★★★★★

Certificate: No certification exam taken

Motivation

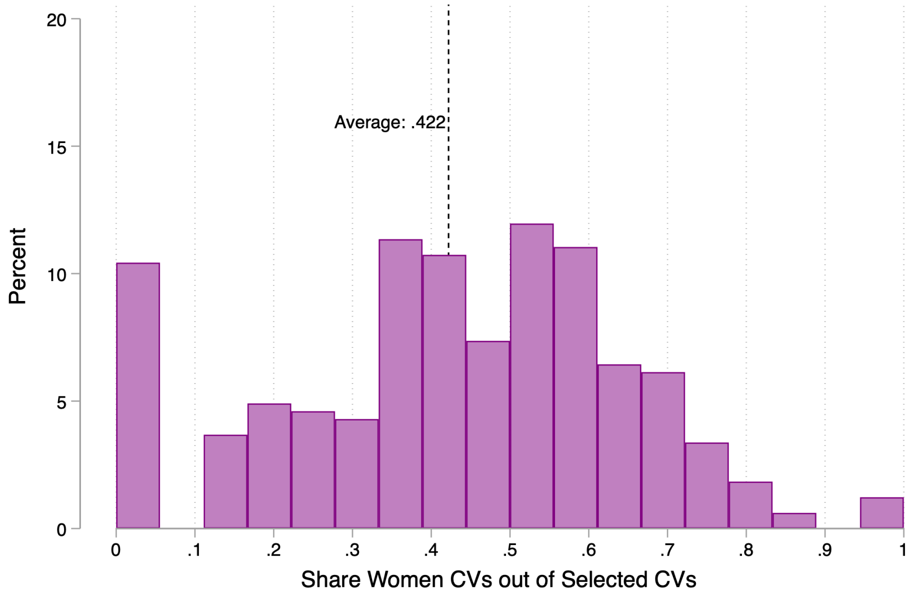
I want to learn practical skills that I can apply in life.

References

Please, call the training center manager Jamila M at 774

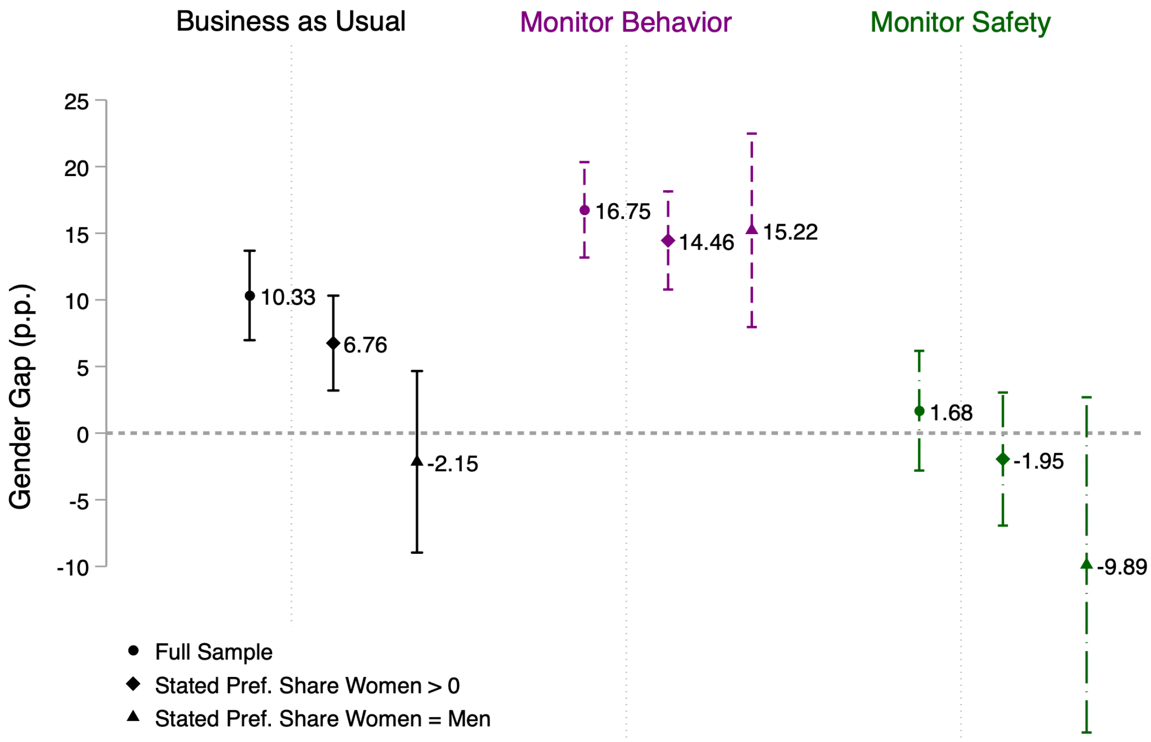
Notes: Examples of hypothetical profile 1 in both female and male versions and for high and low ability levels. In total, there are 24 profiles, with four variations for each profile that reflect the 2×2 design based on gender and ability.

Figure 4: Demand for Female Workers in the Experiment Under Business as Usual



Notes: Data from the main experiment, restricted to the 326 employers in the Pure Control group (PC), comprising 7,365 hiring decisions. The figure shows the distribution of the share of female candidates among those selected for a meeting, with an average female share of 42.2%.

Figure 5: Hiring Gender Gap by Monitoring Arm and Preferences for Workforce Gender-Mix

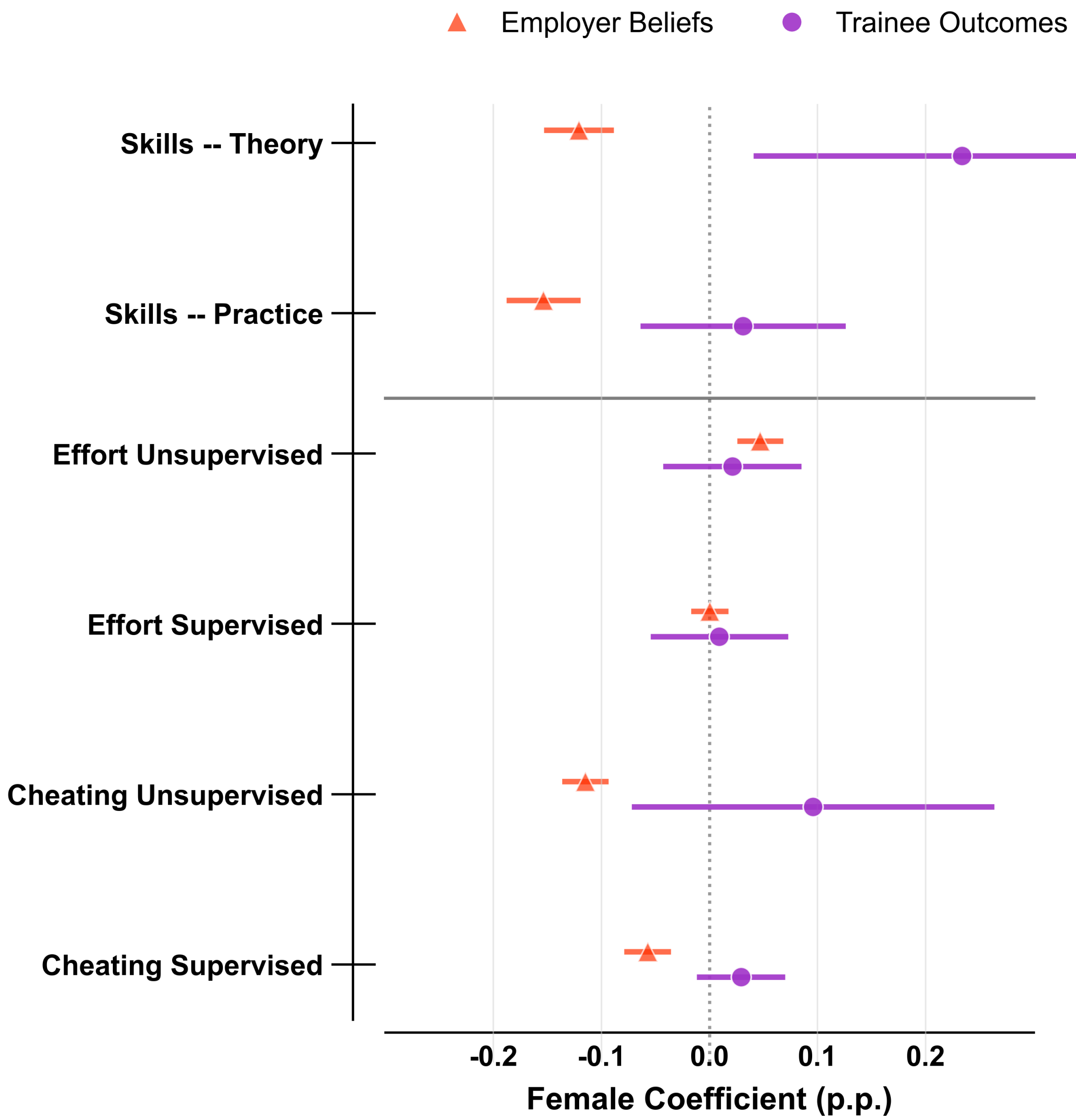


Notes: Data from the experiment. Point estimates and 95% confidence intervals for the gender gap against women are estimated using the regression in Equation 2 and reported by treatment arm. The figure presents results across three levels of employers' stated preferences for workforce gender mix. Circles denote estimates for the full sample ($N = 20{,}726$). Squares denote estimates for employers with any stated preferences for workforce gender mix ($N = 17{,}748$); these are obtained from a regression that includes an indicator for having any stated preferences, fully interacted with candidate gender and the randomly assigned monitoring condition. Triangles denote estimates from the same specification using an indicator for employers in the top quintile of stated preferences (i.e., 40–60% female workers; $N = 3{,}031$).

Figure 6: Gender Gap by Low- and High-Skill Signal Profiles

Notes: Data from the experiment. The figure reports coefficients and 95% confidence intervals from the regression in Equation 3. Circles denote the gender gap against low-skill women and squares the gap against high-skill women. Skill signals proxy technical ability using VTI GPA rankings: the low-skill signal corresponds to the median GPA in the VTI class and is shown as three out of five stars on the profile, while the high-skill signal corresponds to top-of-class GPA scorers and is shown as five out of five stars.

Figure 7: Gender Gap in Skills and Trustworthiness



Notes: Data from the baseline employer survey and the trainee survey. The figure plots the coefficient on gender for employers' beliefs and for actual trainee behavior. Practical technical skills are self-assessed following Alfonsi et al. (2020); theoretical skills are elicited via an exam. The Cheating task is described in Section 5.2.

Tables

Table 1: Descriptive Statistics by Sector

	All	Mechanics		Carpentry		Welding	
	Mean	Mean	[Q25;Q75]	Mean	[Q25;Q75]	Mean	[Q25;Q75]
***Panel A:* Firms**							
Workers (N)	8.34	15.16	[7 ; 20]	4.55	[2.5 ; 5]	5	[3 ; 6]
Female workers (N)	0.14	0.13	[0 ; 0]	0.19	[0 ; 0]	0.09	[0 ; 0]
Trainees (N)	2.64	5.4	[2 ; 6]	0.97	[0 ; 2]	1.41	[0 ; 2]
Family workers (share)	0.14	0.07	[0 ; 0.12]	0.16	[0 ; 0.25]	0.2	[0 ; 0.33]
Age (years)	10.73	12.11	[6 ; 16]	10.01	[5 ; 14]	9.99	[5 ; 13]
Firm capital (USD)	1215.0	1594.4	[0.04 ; 1852]	933.4	[0.04 ; 1037]	1079.1	[0.04 ; 1481]
Firm revenue (USD)	718.1	589.3	[0.04 ; 926]	705.0	[0.04 ; 1111]	874.1	[0.04 ; 1407]
Firm profits (USD)	286.0	246.2	[0.04 ; 370]	276.8	[0.04 ; 370]	339.4	[0.04 ; 556]
Customers (N/day)	11.76	6.44	[4 ; 8]	14.37	[6 ; 20]	14.76	[5 ; 20]
Any stealing	0.83	0.9	[1 ; 1]	0.81	[1 ; 1]	0.77	[1 ; 1]
Monitor constrained	0.9	0.89	[1 ; 1]	0.93	[1 ; 1]	0.89	[1 ; 1]
***Panel B:* Employers**							
Male	0.96	0.96	[1 ; 1]	0.96	[1 ; 1]	0.97	[1 ; 1]
Age (years)	37.08	40.08	[32 ; 47]	35.78	[28 ; 42]	35.24	[29 ; 40]
Experience (years)	15.21	18.66	[10 ; 25]	13.19	[6 ; 19]	13.6	[8 ; 20]
Higher education	0.63	0.64	[0 ; 1]	0.62	[0 ; 1]	0.65	[0 ; 1]
VTI trained	0.31	0.37	[0 ; 1]	0.27	[0 ; 1]	0.29	[0 ; 1]
Hours worked (per day)	10.44	10.47	[10 ; 12]	10.48	[9 ; 12]	10.38	[10 ; 11]
Hours monitor (per day)	2.06	2.03	[1 ; 2]	2.1	[1 ; 2]	2.07	[1 ; 2]
Observations	921	318		300		303	

Notes: The table reports summary statistics for the full sample, with means and interquartile ranges by sector. Panel A presents firm characteristics and Panel B employer characteristics. Monetary values are in USD. *Family workers (share)* is the share of workers who are family members or close friends. *Any stealing*, *Monitoring constrained*, *Higher education*, and *VTI trained* are binary indicators for having experienced worker theft, desiring more monitoring, completing secondary education, and having received VTI training, respectively.

Table 2: Balance Table

	PC (1)		MB (2)		MS (3)		p-value	p-value	N
	Mean	SD	Mean	SD	Mean	SD	(1)-(2)	(1)-(3)	
***Panel A:* Firms**									
Workers (N)	8.24	(8.96)	8.67	(8.66)	7.95	(9.14)	0.085	0.502	903
Female workers (N)	0.19	(0.55)	0.21	(0.56)	0.17	(0.48)	0.464	0.831	906
Trainees (N)	2.52	(3.43)	2.57	(3.25)	2.81	(5.48)	0.381	0.097	906
Family workers (share)	0.15	(0.25)	0.15	(0.23)	0.13	(0.22)	0.527	0.115	898
Age (years)	10.62	(7.82)	11.27	(8.11)	10.35	(7.23)	0.244	0.671	906
Firm capital (USD)	1245	(2776.4)	1437	(3547.9)	974.7	(1421.8)	0.220	0.136	866
Firm revenue (USD)	649.2	(861.9)	699.9	(885.8)	805.6	(1055.4)	0.556	0.122	861
Firm profits (USD)	259.8	(396.4)	279.8	(405.2)	317.1	(468.7)	0.634	0.187	865
Customers (N/day)	11.9	(17.65)	12.03	(13.2)	11.1	(11.87)	0.810	0.109	906
Any stealing	0.83	(0.38)	0.85	(0.36)	0.81	(0.39)	0.667	0.528	906
Monitoring constrained	0.9	(0.31)	0.92	(0.28)	0.9	(0.3)	0.395	0.930	906
***Panel B:* Employers**									
Male	0.97	(0.17)	0.95	(0.23)	0.98	(0.14)	0.115	0.387	906
Age (years)	37.12	(9.84)	36.47	(9.25)	37.26	(10.26)	0.472	0.620	903
Experience (years)	15.12	(8.96)	15.38	(9.11)	14.81	(9.01)	0.580	0.719	904
Higher education	0.62	(0.49)	0.62	(0.49)	0.66	(0.47)	0.966	0.491	906
VTI trained	0.27	(0.45)	0.34	(0.47)	0.32	(0.47)	0.058	0.159	906
Hours worked (per day)	10.55	(1.54)	10.45	(1.7)	10.35	(1.82)	0.592	0.237	906
Hours monitor (per day)	2.08	(1.58)	2	(1.29)	2.08	(1.52)	0.459	0.970	902

Notes: The table presents baseline balance for the main experimental sample. Columns (1), (2), and (3) report means by treatment arm—Pure Control (PC), Monitoring Behavior (MB), and Monitoring Safety (MS)—with standard deviations shown in parentheses in adjacent columns. Columns (1)–(2) and (1)–(3) report p-values from regressions of each outcome on a treatment indicator, including strata fixed effects and robust standard errors. The final column reports the sample size used in each balance test. Some observations are missing due to respondent nonresponse. Significance levels are based on randomization inference p-values: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Hiring Gender Gap in Pure Control Arm and by Stated Preferences

Meet (0-1)	(1) Main Sample	(2) Gender-Mix Pref. > 0	Stated Preferences for Workforce Gender Mix (3) Bottom QT	(4) 2nd QT	(5) 3rd QT	(6) 4th QT	(7) Top QT
Female	-0.103***	-0.077***	-0.233***	-0.153***	-0.098***	-0.083***	0.026
	[0.016]	[0.015]	[0.050]	[0.038]	[0.026]	[0.027]	[0.032]
Observations	7,365	6,206	1,159	870	2,340	1,764	1,232
R-squared	0.123	0.123	0.200	0.190	0.149	0.129	0.137
Control mean	0.522	0.517	0.552	0.497	0.523	0.535	0.491

Notes: The table reports treatment effects from regression model 1 for employers in the Pure Control (PC) group, overall and by quintiles of employers' preferred workforce gender composition. Column (1) presents results for the full PC sample; column (2) excludes employers who prefer an all-male workforce. Columns (3)–(7) restrict the sample to each quintile of the preferred gender composition distribution. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate to hire them on probation. *Female* is an indicator for female candidates. Standard errors, reported in brackets, are clustered at the respondent and profile levels. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Hiring Gender Gap Across Treatment Arms and by Stated Preferences

			Stated Preferences for Workforce Gender Mix				
Meet (0-1)	(1) Main Sample	(2) Gender-Mix Pref. > 0	(3) Bottom QT	(4) 2nd QT	(5) 3rd QT	(6) 4th QT	(7) Top QT
Female	-0.103***	-0.078***	-0.235***	-0.158***	-0.099***	-0.087***	0.029
	[0.016]	[0.016]	[0.052]	[0.039]	[0.026]	[0.028]	[0.034]
Monitor Behavior	0.052***	0.061***	-0.014	0.038	0.052*	0.060**	0.104***
	[0.014]	[0.015]	[0.040]	[0.037]	[0.027]	[0.026]	[0.034]
Monitor Safety	-0.013	-0.017	-0.000	-0.061*	-0.016	-0.020	-0.005
	[0.015]	[0.017]	[0.047]	[0.035]	[0.026]	[0.027]	[0.044]
Female x Monitor Behavior	-0.064***	-0.077***	0.003	-0.050	-0.061	-0.040	-0.178***
	[0.022]	[0.023]	[0.071]	[0.061]	[0.038]	[0.044]	[0.047]
Female x Monitor Safety	0.086***	0.091***	0.026	0.137**	0.104**	0.091*	0.070
	[0.028]	[0.028]	[0.089]	[0.054]	[0.044]	[0.045]	[0.073]
Observations	20,725	17,748	2,977	3,002	6,471	5,244	3,031
R-squared	0.127	0.126	0.177	0.175	0.147	0.119	0.124
Control Mean	0.522	0.517	0.552	0.497	0.523	0.535	0.491

Notes: The table reports the treatment effects for the main sample, and by quintiles distribution of stated preferred gender workforce composition of the employer. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintiles distribution of preferred gender workforce composition. The dependent variable is *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate to hire them on probation. *Female* is a binary variable indicating whether the profile corresponds to a woman. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A Appendix

A.1 Differences From Preregistration

We note three main deviations from the first pre-registration. First, the second primary outcome was not recorded as intended due to a coding error in the survey. The variable *Offer*—elicited using the question, "How likely would you be to offer this worker a position as a mechanic at your firm? Please rate on a scale from 0 to 10, where 0 is very unlikely and 10 is very likely"—was intended to be a primary outcome but was not recorded in the first wave; we therefore chose not to collect it in the second wave.

Second, we expanded the study to include two additional sectors—carpentry and welding—in addition to mechanics. As preregistered, we planned to expand data collection to additional sectors if power proved insufficient. We observed substantial heterogeneity in stated preferences for hiring women, and the additional sectors provide sufficient power to study treatment effect heterogeneity by stated preferences.

Third, because the number of firms per new sector in the second wave was uncertain, we added a *within-subject* component. After completing the initial 24 evaluations under the main treatment condition, each employer rated 12 additional profiles under a second monitoring condition.

Fourth, although the safety audit arm was preregistered as an active control or benchmark, we report its results as an additional treatment of interest given their policy relevance.

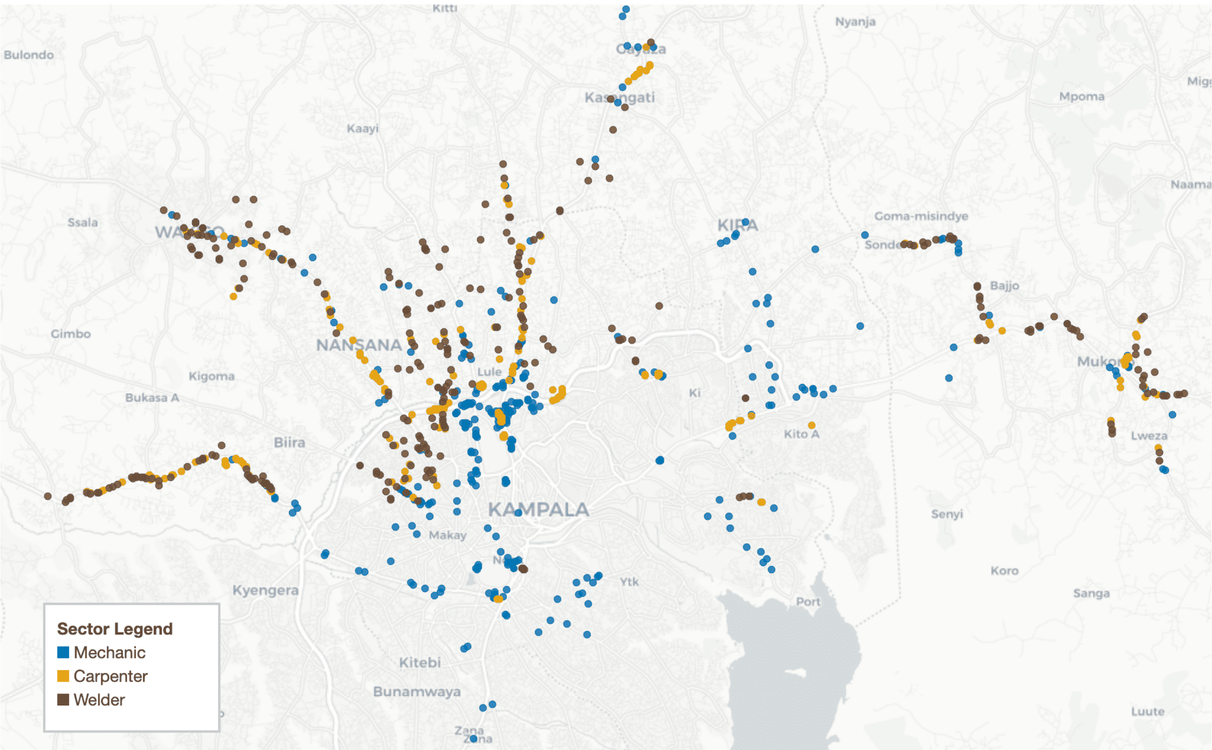
We amended the preregistration before expanding the sample to reflect these changes (except the last one) and to document the motivations.

A.2 Within-Subject Design Results

The within-subject design is not our primary specification, but it provides greater statistical power. The results, summarized in Table A10, are consistent in sign and similar in magnitude to those from the between-subjects design. For example, providing an employer with monitoring support reduces demand for female workers and increases the gender gap in hiring against women by 100.1% (12.2 percentage points increase, p-value 0.000) compared to a business-as-usual condition (Table A10, column (1)).

A.3 Appendix Figures

Figure A1: Firm's Location



Notes: This figure reports the locations of firms in the study sample in the metropolitan area of Kampala, Uganda. For privacy, locations accuracy has been reduced by showing a random positions within 100m radius of the actual location.

Figure A2: Firm and Vocational Training Environment

(a) Firm Layout

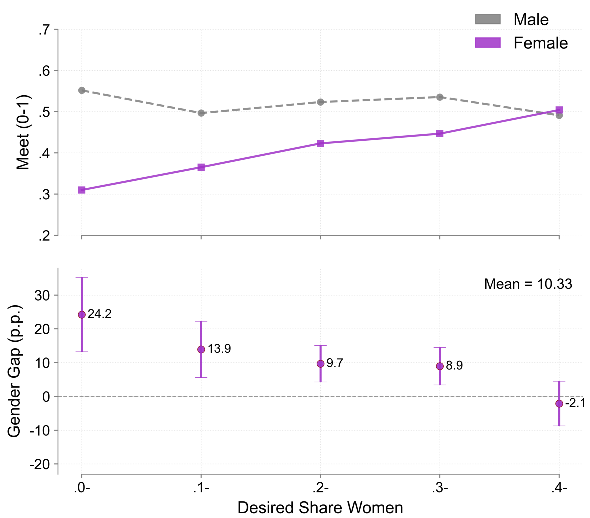
(b) Practical VTI Class

(c) Internship

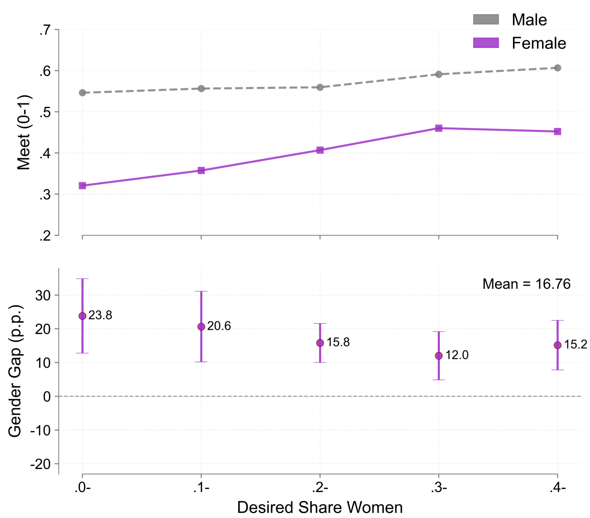
Photo credit: © Mariajose Silva-Vargas.

Figure A3: Hiring Gender Gap by Treatment Arm and by Stated Preferences for Hiring Women (Raw Data)

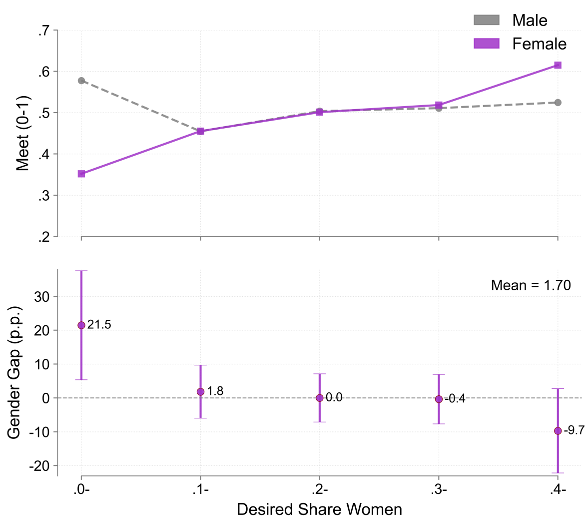
(a) Pure Control (PC)



(b) Monitoring Behavior (MB)



(c) Monitoring Safety (MS)



Notes: The figures show the heterogeneity in the gender gap by stated preferences for workforce gender mix, by treatment arm. Panel (a) presents descriptive statistics for the Pure Control group, while panels (b) and (c) present the corresponding statistics for the Monitoring Behavior and Monitoring Safety groups, respectively. The top graph of each panel presents the raw data average of the employer's interest in meeting the candidate by profile gender by the share of female workers preferred at baseline, while the bottom graph displays the heterogeneous treatment effect of a profile being assigned to a female candidate, estimated with our main specification from Equation 2. Confidence intervals are at the 95% level.

Figure A4: Theory Technical Exam — An Example (Motor Mechanics)

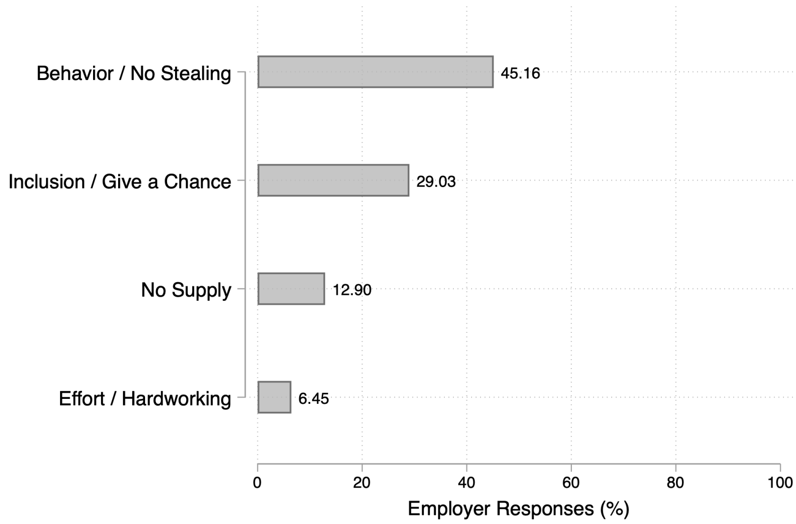
FULL NAME:______________________ ☐ MAN

ID:____________________ ☐ WOMAN

MOTOR-MECHANICS

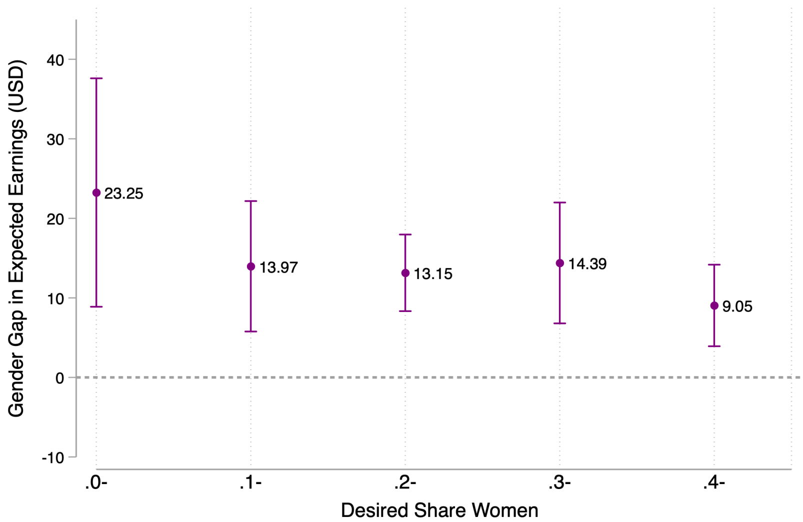
	Question	Answers
1	*multiple-choice* What are you advised to do when servicing the engine by changing oil?	A. Top up lubricating oil B. Replace oil filter C. Over hand engine D. Over hand cylinder head
2	*multiple-choice* What immediate remedy can you give to a vehicle with a problem of excessive tyre wear in the center more than other parts?	A. Increase tyre pressure B. Reduce tyre pressure C. Inflate pressure D. Remove the vehicle tire
3	*Multiple-choice* If a customer reports to you that his/her vehicle charging system works at lower rate, how can you help him?	A. Replacing the charging system B. Adjusting the alternator tension C. Replacing alternator housing D. Renewing wire insulator
4	*Multiple-choice* Which of the following set of systems or component call for mechanical adjustment during general vehicle service?	A. Tyres, cooling system, master cylinder B. Break shoes, alternator, and valve clearance C. Distributor, radiator, propeller shaft D. Tank, crank shaft, Turbo charger
5	*Multiple-choice* What solution would you give a customer with a vehicle engine producing blue smoke?	A. Top up lubricant B. Time the engine C. Replace piston rings D. Remove carbon deposits
6	*Matching* What should you do to stop the following vehicle troubles?	1. Battery over charging — A. Leaking fuel tank 2. Engine over heating — B. Renew regulator 3. Lubricant leakage — C. Reduce oil to the correct level 4. Smoke in exhaust — D. Renew piston rings 5. Engine fails to start — E. Charge the battery ____ ; ____ ; ____ ; ____ ; ____ ; ____ .
7	*Order* When changing engine oil, in which order should you perform the following steps?	A. Drain oil through drain plug B. Remove oil filter cup C. Run engine to check leaks D. Fill new oil through filler cup to level E. Remove oil filter F. Warm up the engine 1____ ; 2____ ; 3____ ; 4____ ; 5____ ; 6____ .

Figure A5: Motivation for Selecting Female Workers in Experiment if No Stated Preferences



Notes: Data from employers who express a preference for male workers only and currently do not employ any women, yet selected at least one female CV in the experiment ($N = 32$). We ask an open-ended question: "Can you explain why you were choosing women in the exercise with the profiles before?". Responses were re-coded into standardized categories, with 9.68% falling under "Other". "No Supply" indicates question misunderstanding (e.g.: "There are no women workers that have come to ask for jobs").

Figure A6: Beliefs About Gender Gap in Labor Market Outside Options



Note: Data from the experiment, focusing on employers in the Pure Control group. The figure reports the predicted gender gap in monthly earnings (USD) of the candidates a year from the hire by stated preferences for workforce gender mix. The average gender gap is USD 14, with men expected to earn an average of USD 117.9 monthly (12.3%).

A.4 Appendix Tables

Table A1: Attention Checks and Attrition

	Monitor Treat		Female Treat
	(1) PC vs. MB Sample	(2) PC vs. MS Sample	(3) PC Sample
Main Sample	0.140	-0.092	-0.032
	[0.125]	[0.058]	[0.022]
Observations	14,954	14,916	7,812
R-squared	0.001	0.002	0.000
p-value	0.264	0.113	0.155

Notes: The table reports the tests for differential attrition by monitoring treatment assignments due to attention checks. The dependent variable is *Main Sample*, a binary indicator equal to 1 if the evaluation of a given CV is included in the main sample. An evaluation is not included in the main sample if all the preregistered attention checks are not met: the respondent understood the IRR exercise, and a given block of eight profiles has some variation in the evaluations. The dependent variables are a binary indicator equal to 1 if the respondent is randomly assigned to: Monitoring Behavior (MB) arm in column (1); Monitoring Safety (MS) arm in column (2), and Female profile in column (3). Samples used to estimate the regressions are restricted to Pure Control (PC) and MB, PC and MS, and PC only in columns (1)-(3), respectively. Standard errors in brackets are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Hiring Gender Gap by High/Low Ability Signal and by Treatment Arm

			Stated Preferences for Workforce Gender Mix				
	(1) Main Sample	(2) Gender-Mix Pref. > 0	(3) Bottom QT	(4) 2nd QT	(5) 3rd QT	(6) 4th QT	(7) Top QT
Panel A: **Pure Control**							
Female	-0.087***	-0.058**	-0.233***	-0.148***	-0.048	-0.083**	0.034
	[0.022]	[0.021]	[0.052]	[0.051]	[0.031]	[0.037]	[0.039]
Female x High Ability	-0.033	-0.042	0.004	-0.008	-0.103***	-0.002	-0.016
	[0.032]	[0.029]	[0.062]	[0.061]	[0.033]	[0.053]	[0.052]
High Ability	0.118***	0.121***	0.105**	0.121**	0.141***	0.117***	0.093*
	[0.018]	[0.019]	[0.044]	[0.053]	[0.025]	[0.040]	[0.047]
Observations	7,365	6,206	1,159	870	2,340	1,764	1,232
R-squared	0.134	0.134	0.212	0.204	0.160	0.142	0.144
Control mean	0.464	0.459	0.491	0.445	0.448	0.481	0.455
Panel B: **Monitor Behavior**							
Female	-0.153***	-0.148***	-0.169***	-0.249***	-0.147***	-0.094**	-0.140***
	[0.024]	[0.025]	[0.053]	[0.064]	[0.036]	[0.038]	[0.050]
Female x High Ability	-0.028	-0.015	-0.117**	0.077	-0.021	-0.063	-0.021
	[0.023]	[0.023]	[0.048]	[0.057]	[0.037]	[0.048]	[0.054]
High Ability	0.111***	0.107***	0.147***	0.010	0.099***	0.190***	0.103**
	[0.021]	[0.020]	[0.040]	[0.044]	[0.032]	[0.037]	[0.046]
Observations	6,830	5,847	983	1,070	2,134	1,620	1,023
R-squared	0.148	0.150	0.205	0.190	0.185	0.132	0.182
Control mean	0.516	0.523	0.475	0.555	0.510	0.499	0.556
Panel C: **Monitor Safety**							
Female	-0.005	0.027	-0.228**	-0.029	0.000	0.046	0.109
	[0.025]	[0.025]	[0.091]	[0.042]	[0.040]	[0.041]	[0.071]
Female x High Ability	-0.024	-0.029	-0.011	0.008	0.005	-0.084**	-0.019
	[0.025]	[0.027]	[0.047]	[0.037]	[0.043]	[0.040]	[0.068]
High Ability	0.081***	0.090***	0.058	0.017	0.084*	0.131***	0.070
	[0.016]	[0.016]	[0.047]	[0.023]	[0.044]	[0.022]	[0.055]
Observations	6,530	5,695	835	1,062	1,997	1,860	776
R-squared	0.102	0.103	0.219	0.178	0.111	0.116	0.103
Control mean	0.468	0.452	0.556	0.437	0.454	0.446	0.484

Notes: The table reports the treatment effects from estimation regression models 3 separately for the sample of respondents in Pure Control (PC) arm (Panel A), the Monitoring Behavior (MB) arm (Panel B) and in the Monitoring Safety (MS) arm (Panel C), and by quintiles distribution of preferred gender workforce composition of the employer. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintiles distribution of preferred gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate to hire them on probation. *Female* is a binary variable indicating whether the profile corresponds to a female candidate. *High-Ability* is a dummy indicating whether the profile has a top GPA score (5/5) within their vocational training program. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Hiring Gender Gap by Treatment Arm and High/Low Ability Index

	(1) Main Sample	(2) Gender-Mix Pref. > 0	(3) Bottom QT	(4) 2nd QT	(5) 3rd QT	(6) 4th QT	(7) Top QT
			Stated Preferences for Workforce Gender Mix				
***Panel A:* Pure Control**							
Female	-0.104***	-0.078***	-0.231***	-0.138***	-0.100***	-0.090***	0.026
	[0.016]	[0.016]	[0.050]	[0.039]	[0.025]	[0.026]	[0.033]
Female x Ability Index	-0.009	-0.010	-0.010	-0.029	0.021	-0.043*	0.000
	[0.011]	[0.012]	[0.010]	[0.032]	[0.017]	[0.024]	[0.030]
Ability Index	0.133***	0.131***	0.141***	0.170***	0.104***	0.167***	0.107**
	[0.017]	[0.018]	[0.044]	[0.053]	[0.033]	[0.038]	[0.046]
Observations	7,365	6,206	1,159	894	2,340	1,740	1,232
R-squared	0.133	0.133	0.212	0.195	0.157	0.146	0.144
Control mean	0.478	0.480	0.466	0.460	0.503	0.477	0.451
***Panel B:* Monitor Behavior**							
Female	-0.168***	-0.155***	-0.228***	-0.210***	-0.158***	-0.126***	-0.150***
	[0.017]	[0.018]	[0.051]	[0.050]	[0.028]	[0.033]	[0.038]
Female x Ability Index	0.001	0.010	-0.038*	0.022	0.002	-0.019	0.045
	[0.013]	[0.014]	[0.022]	[0.038]	[0.024]	[0.019]	[0.029]
Ability Index	0.121***	0.121***	0.128***	0.050	0.110***	0.210***	0.094
	[0.020]	[0.021]	[0.045]	[0.048]	[0.027]	[0.042]	[0.057]
Observations	6,830	5,847	983	1,070	2,134	1,620	1,023
R-squared	0.148	0.150	0.204	0.189	0.185	0.131	0.184
Control mean	0.538	0.545	0.500	0.540	0.502	0.539	0.648
***Panel C:* Monitor Safety**							
Female	-0.017	0.013	-0.234***	-0.025	0.002	0.004	0.099
	[0.022]	[0.021]	[0.079]	[0.038]	[0.035]	[0.035]	[0.060]
Female x Ability Index	-0.017	-0.019	-0.042	-0.009	-0.004	-0.027	-0.027
	[0.012]	[0.013]	[0.030]	[0.018]	[0.023]	[0.023]	[0.033]
Ability Index	0.095***	0.104***	0.085	0.030	0.110***	0.125***	0.090
	[0.016]	[0.015]	[0.058]	[0.035]	[0.034]	[0.026]	[0.063]
Observations	6,530	5,695	835	1,062	1,997	1,860	776
R-squared	0.102	0.104	0.221	0.178	0.111	0.115	0.104
Control mean	0.451	0.447	0.480	0.382	0.458	0.458	0.485

Notes: The table reports the treatment effects from estimation regression models 3 for the sample of respondents in the Monitoring Behavior (MB) arm (Panel A) and in the Monitoring Safety (MS) arm (Panel B), and by quintiles distribution of preferred gender workforce composition of the employer. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintiles distribution of preferred gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate to hire them on probation. *Female* is a binary variable indicating whether the profile corresponds to a female candidate. *Ability Index* is a standardized index that combines all high-quality trainee characteristics in the CVs with equal weights: completion of secondary education, a Directorate of Industrial Training (DIT) certification, a 12- (vs. 6-month) study period, and a high GPA. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Hiring Gender Gap by High and Low Ability Signals and Monitoring (Non-parametric, Fully Saturated)

	(1) Hiring Choice
Female	-0.087***
	[0.022]
High Ability	0.118***
	[0.018]
Female x High Ability	-0.033
	[0.032]
Monitor Behavior	0.055***
	[0.018]
Monitor Safety	0.006
	[0.019]
Female x Monitor Behavior	-0.067**
	[0.027]
Female x Monitor Safety	0.082**
	[0.030]
High Ability x Monitor Behavior	-0.007
	[0.021]
High Ability x Monitor Safety	-0.038*
	[0.022]
Female x High Ability x Monitor Behavior	0.005
	[0.030]
Female x High Ability x Monitor Safety	0.009
	[0.029]
Observations	20,726
R-squared	0.126
Mean: Male, Low Ability in PC	0.464

Notes: The table reports the treatment effects for the main sample. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate to hire them on probation. *Female* is a binary variable indicating whether the profile corresponds to a woman. *High-Ability* is a dummy indicating whether the profile has a top GPA score (5/5) within their vocational training program. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. The control mean is that of Male candidates with Low Ability (i.e. median GPA score within their program) in *Pure Control* arm. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Hiring Gender Gap by Treatment Arm: Secondary Outcomes

	(1) Quality	(2) Trust-worthiness	(3) Effort	(4) Expected Earnings
Female	-0.169***	0.058**	-0.209***	-12.395***
	[0.026]	[0.026]	[0.028]	[1.578]
Monitor Behavior	-0.013	-0.034	-0.012	2.267
	[0.041]	[0.049]	[0.042]	[3.193]
Monitor Safety	-0.030	-0.035	-0.016	-1.611
	[0.048]	[0.053]	[0.044]	[3.200]
Female x Monitor Behavior	-0.043	-0.012	-0.054	-0.933
	[0.028]	[0.035]	[0.037]	[2.394]
Female x Monitor Safety	0.034	0.059	0.019	-0.556
	[0.047]	[0.042]	[0.044]	[2.414]
Observations	20,724	20,613	20,684	20,725
R-squared	0.226	0.083	0.173	0.196
PC Mean	0.085	-0.028	0.106	114.913

Notes: The table reports the treatment effects for the main sample, by pre-registered secondary outcomes. Columns (1)-(3) reports standardized outcomes relative to PC group. Column (1) shows estimates for the perceived quality and skills of candidates: "Based on your first impression, how would you rate the worker's skills and work quality? Please rate on a scale from 0 to 10, where 0 is very low quality and 10 is very high quality.". Column (2) shows estimates for the perceived trustworthiness of candidates: "Based on your first impression, how would you rate the workers' behavior (trustworthiness and honesty)? Please rate on 0 to 10, where 0 is not at all trustworthy/honest and 10 is very trustworthy/honest.". Column (3) shows estimates for the perceived effort of candidates: "Based on your first impression, how likely is this worker to put in effort (hardworking, not lazy, concentrated, punctual)? Please rate on a scale from 0 to 10, where 0 is very unlikely and 10 is very likely.". Column (4) shows estimates for the expected monthly earnings of the candidates: "What is your best guess of the monthly earnings of this worker a year from now?". Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Effect of Audit Visits on Number of Selected Profiles

	(1) Meet (0-1)
Monitor Behavior	0.019*
	[0.011]
Monitor Safety	0.030**
	[0.011]
Observations	20,725
R-squared	0.114
Control Mean	0.470

Notes: The table reports the treatment effects for the main sample. We estimate the following regression model: $\text{Meet}_{ijs} = \beta_0 + \beta_1 \text{MB}j + \beta_2 \text{MS}j + \delta_i + \sigma_s + u_{ijs}$. *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate for a probation period at the firm. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to behavior audits. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. The control mean corresponds to the *Pure Control* arm. Profile and strata fixed effects are included (not reported). Standard errors are clustered at the respondent and profile levels and reported in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Hiring Gender Gap by Treatment Arm and by Resume Characteristics

Meet (0-1)	(1) Age $\geq$ Median	(2) Married	(3) Second. Educ.	(4) English	(5) Trained 1+ year	(6) DIT Certified
CHRS	-0.094***	-0.209***	0.007	0.067***	0.209***	0.145***
	[0.012]	[0.011]	[0.012]	[0.018]	[0.011]	[0.013]
Monitor Behavior	0.012	0.009	0.025*	0.016*	0.029**	0.029***
	[0.014]	[0.012]	[0.015]	[0.009]	[0.012]	[0.010]
Monitor Safety	0.037***	0.022*	0.034**	0.033***	0.040***	0.038***
	[0.014]	[0.012]	[0.015]	[0.009]	[0.012]	[0.010]
CHRS x Monitor Behavior	0.010	0.019	-0.010	0.016	-0.019	-0.035*
	[0.017]	[0.016]	[0.018]	[0.025]	[0.016]	[0.018]
CHRS x Monitor Safety	-0.008	0.019	-0.003	-0.009	-0.019	-0.021
	[0.017]	[0.017]	[0.018]	[0.026]	[0.017]	[0.018]
Observations	20,725	20,725	20,725	20,725	20,725	20,725
Control Mean	0.470	0.574	0.466	0.462	0.365	0.428

Notes: The table reports the treatment effects for the main sample, by resume characteristics (CHRS). Estimates are shown in Columns (1)–(6), respectively, for whether the candidate: is above the median age (22.5 years), is married, has attended secondary education, speaks English, have been attending VTI one- or two-year training relative to 6-month training, holds a Directorate of Industrial Training (DIT) certification. Strata fixed effects are included (not reported in the table). Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Hiring Gender Gap by Treatment Arm – Heterogeneous Treatment Effects (Preregistered)

	Gender-Mix Pref. > 0		Employer Gender		VTI Search	
Meet (0-1)	(1) Low	(2) High	(3) Female	(4) Male	(5) Yes	(6) No
Female	-0.146***	-0.041*	0.001	-0.105***	-0.086***	-0.174***
	[0.021]	[0.022]	[0.111]	[0.017]	[0.018]	[0.042]
Monitor Behavior	0.035*	0.080***	0.149	0.049***	0.067***	-0.027
	[0.019]	[0.020]	[0.100]	[0.014]	[0.017]	[0.032]
Monitor Safety	-0.021	-0.006	0.280***	-0.018	-0.014	-0.008
	[0.018]	[0.024]	[0.075]	[0.016]	[0.017]	[0.041]
Female x Monitor Behavior	-0.043	-0.092***	-0.206	-0.061**	-0.086***	0.027
	[0.031]	[0.032]	[0.138]	[0.023]	[0.024]	[0.065]
Female x Monitor Safety	0.098***	0.075*	-0.218*	0.092***	0.081**	0.096
	[0.033]	[0.041]	[0.125]	[0.028]	[0.032]	[0.069]
Observations	12,450	8,275	727	19,998	17,108	3,545
R-squared	0.147	0.112	0.173	0.130	0.127	0.172
Control Mean	0.528	0.520	0.465	0.524	0.513	0.566

Notes: The table reports the treatment effects for the main sample, by pre-registered heterogeneity dimensions: Columns (1)-(2) show estimates for above- (high) and below-median (low) optimal gender composition of the workforce; Columns (3)-(4) split the sample by the employers' gender; Columns (5)-(6) split the sample by whether employers typically hire from vocational training institutes or not. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants to meet the candidate to hire on probation. *Female* is a binary variable indicating whether the profile corresponds to a woman. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. Profile and strata fixed effects are included. Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Gender Gap by Treatment Assignment and Lin. Interaction with Gender Pref.

Meet (0-1)	(1) Pure Control	(2) Monitor Behavior	(3) Monitor Safety	(4) Main Sample
Female	-0.219***	-0.217***	-0.139***	-0.218***
	[0.034]	[0.033]	[0.044]	[0.034]
Monitor Behavior				-0.002
				[0.027]
Monitor Safety				-0.026
				[0.029]
Gender-Mix	-0.106	0.132*	-0.048	-0.111
	[0.084]	[0.077]	[0.101]	[0.080]
Female x Monitor Behavior				0.002
				[0.046]
Female x Monitor Safety				0.076
				[0.054]
Female x Gender-Mix	0.521***	0.224**	0.564***	0.514***
	[0.121]	[0.105]	[0.165]	[0.122]
Monitor Behavior x Gender-Mix				0.243**
				[0.101]
Monitor Safety x Gender-Mix				0.062
				[0.124]
Female x Monitor Behavior x Gender-Mix				-0.296*
				[0.161]
Female x Monitor Safety x Gender-Mix				0.062
				[0.209]
Observations	7,365	6,830	6,530	20,725
R-squared	0.130	0.144	0.105	0.134
Baseline Mean	0.522	0.572	0.522	0.552

Note: The table reports the treatment effects on the interest of employers to meet the candidates by stated preferences for workforce gender mix of employers. Columns (1)-(3) include all respondents assigned to the Pure Control (PC), Monitoring Behavior (MB), and Monitoring Safety (MS) arms in the between-subject design (main specification). Column (4) includes all respondents in the main sample and estimates a fully interacted model with candidate gender, monitoring treatment, and employer stated preferences for workforce gender mix. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants to meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the resume corresponds to a woman. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. *Gender-Mix* is a variable ranging from 0 to 1 that measures the share of female workers preferred by the employer in their ideal workforce composition. *Baseline Mean* is the average of *Meet* for male resumes evaluation by employers with stated preferences for an all-male workforce in PC. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered at the employer and profile levels (reported in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

Table A10: Hiring Gender Gender by Treatment Arm: Within-Subject Design

			Stated Preferences for Workforce Gender Mix				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meet (0-1)	Main Sample	Gender-Mix Pref. > 0	Bottom QT	2nd QT	3rd QT	4th QT	Top QT
Female	-0.115***	-0.096***	-0.236***	-0.122**	-0.100***	-0.122***	-0.050**
	[0.015]	[0.015]	[0.055]	[0.049]	[0.024]	[0.026]	[0.024]
Monitor Behavior	0.092***	0.103***	0.037	0.076**	0.121***	0.083***	0.106***
	[0.013]	[0.013]	[0.038]	[0.032]	[0.017]	[0.025]	[0.031]
Monitor Safety	-0.046***	-0.045***	-0.018	-0.092***	-0.041*	-0.050**	-0.027
	[0.013]	[0.014]	[0.043]	[0.027]	[0.024]	[0.021]	[0.025]
Female x Monitor Behavior	-0.122***	-0.125***	-0.119*	-0.149**	-0.129***	-0.096**	-0.123***
	[0.020]	[0.020]	[0.064]	[0.059]	[0.029]	[0.039]	[0.036]
Female x Monitor Safety	0.149***	0.163***	-0.019	0.179***	0.163***	0.174***	0.166***
	[0.024]	[0.024]	[0.085]	[0.057]	[0.042]	[0.040]	[0.046]
Observations	20,800	18,409	2,391	2,447	6,521	5,714	3,727
R-squared	0.221	0.214	0.314	0.249	0.237	0.208	0.191
Control Mean	0.529	0.529	0.536	0.522	0.513	0.550	0.526
p-value MB vs. MS	0.000	0.000	0.240	0.000	0.000	0.000	0.000

Note: The table reports the treatment effects for the sample for which we implemented the within-subject randomization. The table also reports the treatment effects by quintile distribution of preferred gender workforce composition of the employer. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintile's distribution of preferred gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate to hire them on probation. *Female* is a binary variable indicating whether the profile corresponds to a woman. *Monitoring Behavior* is a dummy indicating whether the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a dummy indicating whether the firm was randomly assigned to audit visits to monitor trainees' safety. We report results for the primary outcome variable: interest of the employer to meet the candidates to hire them on probation (0-1). Respondent and profile-by-ability fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile-by-ability levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Trainee Outcomes and Employer's Beliefs

Panel A: **Trainees**

	Social Preferences				**Skills**		**Supervision & Honesty Game**				**Dictator Game**	**Trust Game**		**Cooperation Game**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Patience	Altruism	Pos. Recipr.	Trusting	Theory	Practice	Task Completed Unsupervised	Task Completed Supervised	Lying Unsupervised	Lying Supervised	Share	Trust Pl. 1	Reciprocate Pl. 2	Cooperate
Female	0.089	0.067	0.064	0.041	0.234**	0.031	0.021	0.009	0.096	0.029	0.177*	-0.375***	-0.005	0.131
	[0.073]	[0.071]	[0.045]	[0.059]	[0.097]	[0.047]	[0.031]	[0.031]	[0.084]	[0.020]	[0.106]	[0.093]	[0.099]	[0.099]
Observations	171	171	171	171	187	182	182	182	182	182	182	182	182	182
R-squared	0.028	0.038	0.064	0.034	0.125	0.021	0.097	0.106	0.032	0.083	0.133	0.096	0.088	0.059
Control Male	0.696	0.681	0.768	0.450	0.752	0.920	0.989	0.989	0.071	0.018	0.723	0.679	0.875	0.830

Panel B: **Employers**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Female	-0.132***	-0.129**	0.069	0.481***	-0.121***	-0.154***	0.047***	0.000	-0.115***	-0.057***	-0.089***	-0.011	-0.035***	-0.018*
	[0.048]	[0.051]	[0.045]	[0.042]	[0.015]	[0.016]	[0.010]	[0.007]	[0.010]	[0.010]	[0.012]	[0.010]	[0.010]	[0.010]
Observations	919	919	920	921	1,205	1,206	1,842	1,829	1,831	1,826	1,838	1,835	1,829	1,831
R-squared	0.012	0.010	0.046	0.312	0.093	0.120	0.016	0.007	0.077	0.025	0.032	0.002	0.008	0.020
Control Male	0.435	0.449	0.253	0.130	0.580	0.607	0.646	0.881	0.466	0.256	0.573	0.534	0.602	0.571

Notes: The table reports gender differences in trainees' outcomes (Panel A) and the corresponding employer beliefs (Panel B). Columns (1)–(4) present standardized survey measures of social preferences following Falk et al. (2018): patience, altruism, positive reciprocity, and trust, respectively. In Panel A, the coefficient on *Female* captures the difference between female and male trainees. In Panel B, *Female* captures the difference in employer beliefs about female versus male trainees. Columns (5)–(6) present the share of correct answers to sector-specific theory and practical skill questions, respectively. Theory questions are randomly drawn from the technical exam taken by trainees prior to the survey. Practical skills are measured through a self-reported binary indicator on the ability to perform a specific technical task, designed by VTI instructors. The remaining columns shows results of four behavioral games. Columns (7)–(8) report the share of tasks completed in a novel lab-style task — Supervision & Honesty Game — by trainees (Panel A), and employers' beliefs about the same task outcomes under conditions without and with supervision (Panel B). Columns (9)–(10) report the share of lying behavior in the same game. Column (11) reports sharing behavior in a dictator game. Columns (12)–(13) correspond to player 1 (trusting) and player 2 (reciprocating) behavior in a trust game. Column (14) reports decisions to cooperate in a cooperation game. Standard errors are clustered at VTI level in Panel A and at the sector level in Panel B. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.