Job Search and Hiring with Two-sided Limited Information about Workseekers' Skills*

Eliana Carranza[†] Robert Garlick[‡] Kate Orkin[§] Neil Rankin[¶]

July 22, 2019

Abstract

Firms make hiring decisions and workseekers make job search decisions using potentially noisy signals of workseekers' skills and productivity. Noise can distort job search and hiring decisions and lead to lower total employment and earnings. We study the labor market effects of improving information about workseekers' skills to workseekers and/or firms. Certifying workseekers skills and allowing them to share the certification with firms substantially increases employment and earnings. Providing information only to workseekers or only to firms has positive but smaller effects on labor market outcomes, consistent with both workseekers and firms facing frictions. These findings demonstrate quantitatively important information frictions on both sides of the labor market that can be alleviated by improving information about workseekers' skills.

JEL codes: J23, J24, J31, J41, O15, O17

*We are grateful for exceptional research assistance from Wim Louw, Wendy Trott, Shelby Carvalho, Allegra Cockburn, Emmanuel Bakirdjian and Nilmini Herath in South Africa and Lukas Hensel, Svetlana Pimkina, Gabriel Cunha and Laurel Wheeler at our institutions. Thanks to multiple staff at the Harambee Youth Employment Accelerator; the Abdul Latif Jameel Latif Poverty Action Lab Africa Office (J-PAL Africa); and administrative staff at the Universities of Cape Town (UCT), Duke, Oxford, and Stellenbosch, the Centre for Study of African Economies, and the Africa Gender Innovation Lab and Jobs Practice at the World Bank. The project is conducted in collaboration with the World Bank Africa Gender Innovation Lab, Jobs Practice and Jobs Trust Fund, and received funding from the National Science Foundation (# 1824413), Private Enterprise Development in Low Income Countries (# 3024) and the Global Challenges Research Fund Accelerating Adolescent Achievement Hub. Orkin acknowledges support from the Wellspring Philanthropic Fund. This study has been approved by the Commerce Faculty Ethics in Research Committee at UCT, the Duke Institutional Review Board for the Protection of Human Subjects in Non-Medical Research (protocol # D0368) and the Department of Economics Research Ethics Committee at Oxford (protocol # ECONCIA15-055). This study is pre-registered at https://www.socialscienceregistry.org/trials/1631. The authors' views do not necessarily represent the views of the World Bank or the governments of the countries they represent.

[†]World Bank, ecarranza@worldbank.org

[‡]Duke University, robert.garlick@duke.edu

[§]University of Oxford, kate.orkin@bsg.ox.ac.uk

[¶]University of Stellenbosch, neilrankin@sun.ac.za

1 Introduction

Many countries face high and rising unemployment and underemployment, particularly for young people (World Bank, 2018). Wage work is often short-term and insecure (Donovan et al., 2018) and wages grow slowly over time (Lagakos et al., 2018). These challenges occur across both developed and developing countries in Latin America, the Middle East, North Africa, South Europe, and sub-Saharan Africa and have major economic, political and social implications. Periods of unemployment may permanently harm youths' employment trajectories (Altonji et al., 2015; Oreopoulos et al., 2012; García-Pérez et al., 2018; Kahn, 2010) and contribute to crime and political instability (Blattman and Annan, 2016; Freeman, 1999). At the same time, rms in many developing countries have low labor productivity (Hall and Jones, 1999; Bloom et al., 2010) and low growth rates (Hsieh and Olken, 2014; Hsieh and Klenow, 2014). Firms report that they cannot nd and hire suitable workers (Hardy and McCasland, 2017).

Limited information about workseekers' skills can contribute to there challenges by distorting job search decision, hiring decisions, and rm-worker matching. Limited information for rms can reduce the expected marginal revenue product from new hires and reduce wages. Limited information can also reduce rm-level employment when rms face binding minimum wages or ring costs. Limited information for workseekers can distort job search decisions. These search distortions can also change individual-level employment and earnings. For example, overcon dent workseekers might reject job o ers with low wages and hence remain unemployed, while undercon dent workseekers might withdraw from search entirely. Information frictions may be particularly important for young workseekers who cannot use prior work experience to learn about their own skills or certify these skills to prospective employers. Information frictions may also be particularly prevalent when educational quali cations are poor proxies for skill.

We use a series of eld experiments to identify how revelation of information about workseekers' skills changes their beliefs about their abilities and the labor market, their job search, employer responses to their applications, and ultimately their labor market outcomes. We study a labor market urban South Africa relevant to the model we sketch above: formal education is weakly correlated with skills, minimum wages bind and rms face high separation costs. We sample from a population where information frictions are likely: unemployed and underemployed youths from low-income homes with limited post-secondary education, work experience, and access to referral networks. We recruit roughly 7,000 workseekers and assess their skills using six psychometric assessments.

In our rst experiment, we randomly select some workseekers for a skill certi cation intervention that tells them about their assessment results and allows them to signal these results to prospective employers. After three to four months, these workseekers' employment rate rises by 17% (5 percentage points from a 30% counterfactual rate), weekly earnings rise by 34%, and hourly wages rise by 20%. The earnings e ect re ects both higher employment and higher earnings conditional on employment, which we identify using a new decomposition method. The intervention does not increase the labor market return to skill, which may reflect larger information frictions facing lower-skilled workseekers or demand-side heterogeneity, with some firms preferring low-skilled workers for some vacancies.

The results from the first experiment show information frictions exist, but do not show whether firms, workseekers, or both sides of the market face frictions. For example, workseekers with certificates may learn nothing new about their skills, but change search behavior because they expect firms to value the certificates. Or firms may make offers to certified workseekers because they expect workseekers to have better calibrated their wage and progression expectations to their skill levels. We use additional measurement and experiments to separate these hypotheses.

First, we examine workseekers' behavior. We find that workseekers with certifications have more accurate beliefs about their skills, use their certificates in job applications, and have higher expectations about their labor market outcomes. All but the first of these changes is consistent with either workseekers learning from certification or workseekers changing beliefs and behavior because they expect firms will learn from certification. In our second experiment, we randomly select some workseekers for a `private' certification that tells them their assessment results but does not help them signal these results to employers[1]. While public certification is conceptually similar to relatively credible signals like formal education qualifications, private certification is conceptually similar to assessments without signals like skills assessment as part of search counselling. This intervention changes workseekers' beliefs about their skills roughly as much as the `public' certification in the first experiment but does not change their beliefs about their labor market prospects. This intervention has no effect on employment and has smaller effects on weekly earnings and hourly wages than public certification. We interpret this as evidence that workseekers have somewhat limited information about their skills and change behavior in response to information, even when it is unlikely they can reveal it to firms. However, providing them with information without help signalling it to firms has limited effects on their labor market outcomes.

Second, we examine firm responses to information. Results in the certification experiment may arise because of changes in workseekers' decisions about where to apply and may be affected by whether workseekers share their reports. In our thirdongoing experiment, we randomly vary whether firms receive information about workseekers' skills in job applications. We apply to jobs by email on behalf of real workseekers in our sample, using applications they create. We apply to multiple jobs for each workseeker, attaching public skill certificates only to some randomly chosen applications. We submit applications from four workseekers to each job, and also randomize the

[1] The public certificates in the first experiment show the workseeker's name and national identity number and are branded by the prominent social enterprise with which we work. Workseekers receive 20 hard copies of the report, receive an email copy, and can direct firms to a website with more information on the assessments. Workseekers with private certification get one printed unbranded certificate without their name. Both groups have a group session with a psychologist to explain assessment results.

share of applications with reports sent to each vacancy. More skill certificates increase the callback rate at both the vacancy and application level, though the latter effect is imprecisely estimated. We interpret this as evidence that firms face some information frictions, as they respond to information about workseeker skills. We support this conclusion by showing that firms are willing to pay for access to better information in two small incentivized choice experiments. However, attaching skill certificates to applications in this experiment does not necessarily increase the application-level callback rate enough to generate the employment effects observed in the first experiment. This difference might occur because workseekers use the certificates in directed search, while we randomly match workseekers to vacancies in the third experiment. Taken together, these experiments suggest that both sides of the labor market face information frictions and benefit from the revelation of information about workseeker skills. However, information being available to both firms and workers has larger effects than information being revealed to only one side of the market.

We build on a growing literature on information frictions in labor markets. Several panel studies show that the relationship between wages and skills observed by the econometrician but not by firms becomes stronger with tenure particularly for workseekers with less education, consistent with employers learning about skills over time (Altonji and Pierret, 2001; Arcidiaono et al., 2010; Farber and Gibbons, 1996; Kahn and Lange, 2014; MacLeod et al., 2017).[2] Recent experimental work shows that simultaneously providing both workseekers and firms with information about workseekers' skills or past performance can but does not always shift job search and earnings (Abebe et al., 2016; Abel et al., 2019; Bassi and Nansamba, 2017; Pallais, 2014).

We make two major and two smaller contributions to this literature. First, we are the first paper to identify the separate and joint roles of workseeker- and firm-side information frictions.[3] We find that revealing information to only workseekers or only firms can change labor market outcomes. But the labor market effects are larger with two-sided information revelation, potentially because this allows workseekers to better direct search. The location of information frictions is important for welfare and policy design. For example, if information frictions only distort firms' behavior, then the optimal policy intervention may be to provide firms with better information. This might be performed by in-house assessment teams or specialized third-party firms that provide assessment services. But if information frictions also distort workseekers' job search and application decisions, then the gains to firms from such a service will be limited as firms' preferred candidates might not appear in the pool they assess. These factors also help to determine which side of the market, if either, will be willing to pay for access to better information.

Second, we show that alleviating information frictions can substantially increase extensive-margin employment, as well as total earnings and hourly wages. We demonstrate this both empiri-

[2] A related literature finds mixed evidence for differences in labor market outcomes across workers with different qualifications but very similar skills (Alfonsi et al., 2017; Clark and Martorell, 2014).

[3] Abebe et al. (2016), Bassi and Nansamba (2017), and Pallais (2014) all reveal information simultaneously to both workseekers and firms.

cally directly in our certification experiment and indirectly in our audit study and theoretically by showing that information frictions can drive workseekers' value to firms below the minimum wage they can or will offer. The employer learning literature using longitudinal data focuses only on employed workers. Prior experiments alleviating information frictions among inexperienced workseekers have found some effects on earnings and wages but no effects on employment. This difference may occur because we study a labor market with binding minimum wages and firing restrictions, unlike the relatively unregulated markets studies in prior work.[4] The employment margin for inexperienced workseekers is particularly important in the context of information frictions. Without employment experience, workseekers cannot start revealing their skills to the job market through employer learning (Altonji and Pierret, 2001; Farber and Gibbons, 1996). Privately optimal hiring of inexperienced workseekers will be below the social optimum if work experience allows workseekers to reveal their skills to new employers and change jobs, giving firms an incentive to free-ride on other firms' hiring of young workers (Acemoglu and Pischke, 1999; Kahn, 2013). The skill certification we study offers a relatively cheap way to improve transitions into first employment in the many regulated labor markets facing high youth unemployment.[5]

Third, our findings suggest potential extensions to job search coaching and counselling programs. These interventions are common in developed and some developing countries and often give workseekers information about labor market conditions and their suitability for specific jobs (Card et al., 2015). Skill assessment and certification is quick, relatively cheap, and could be incorporated in job coaching and counselling.[6]

Fourth, our findings might help to inform research on algorithm hiring and matching. Recent work shows that algorithmic hiring decisions can raise fill rates and lower turnover (Autor and Scarborough, 2008; Hoffman et al., 2018; Horton, 2017). This shows that firms can more effectively use information available at the time of hiring to make job offer decisions. In contrast, several interventions to match firms and workseekers have not generated large employment or earnings effects (Beam, 2016; Groh et al., 2015; Abebe et al., 2018). We show that both firms and workseekers have limited information about workseekers skills when they make job search and hiring decisions. Providing better information about workseekers skills might enhance the effectiveness of algorithmic

[4] Abebe et al. (2016) and Bassi and Nansamba (2017) find no employment effects of certification-style experiments in respectively Ethiopia and Uganda, though they do find effects on other labor market outcomes.

[5] Two related studies demonstrate that reference letters improve employment for workseekers with prior work experience. Pallais (2014) finds substantial employment and earnings effects of a reference letter experiment on extensive margin employment on a platform for gig-style work, oDesk. Abel et al. (2019) find no average employment effects of a reference letter experiment in South Africa, though they do find employment effects in some subgroups. Both papers provides important evidence on information frictions. But the effects and implications of frictions may differ for inexperienced workseekers without reference letters, for whom there is much less information available.

[6] Other work demonstrates that low-cost additions can enchance the effectiveness of active labor market programs. Belot et al. (2018) show that workseekers search more broadly and obtain more interviews when shown broader a wider range of online job adverts. Altmann et al. (2018) find that information on job search strategies improved on labor market outcomes for the long-term unemployed. Wheeler et al. (2019) show training participants in job readiness training to open and use LinkedIn accounts increases employment, potentially by facilitating access to labor market information.

hiring and matching.

Our work also relates to research on network-based job search and hiring. This is often interpreted as a response to labor market information frictions (Ioannides and Loury, 2004). However, referees' incentives may be imperfectly aligned with firms, partly offsetting information gains from network hiring (Beaman and Magruder, 2012; Fafchamps and Moradi, 2015; Heath, 2018). Network-based hiring may also exclude women or increase socioeconomic inequality (Beaman et al., 2018; Witte, 2019). Network-based hiring is common in our sample and in South Africa in general (Magruder, 2010). The largest share of extra jobs in our skill certification experiment are obtained through formal applications or interviews after referrals. This suggests that skill certification and referrals may interact, with skill certification allowing referees to make more accurate or credible referrals.

In Section 2 of the paper we describe the economic environment. We present a conceptual framework showing that limited information about workseeker skills on either side of the market can distort labor market outcomes. We describe the South African labor market and our sample, arguing that the market and the sample match the salient features of the labor market we model. We then describe the skills we assess and the assessment process. In Section 3 we describe the skill certification experiment and report the treatment effects on labor market outcomes. In Section 4, we explore the mechanisms generating effects on labor market outcomes, starting with the effects of certification on workseeker beliefs and job search. We then describe each additional experiment and the treatment effects. The additional experiments are a supply-side-only certification experiment, a demand-side-facing audit study, two small incentivized choice experiments with firms, and another small experiment with workseekers to test if certification works through some mechanism other than information provision.

2 Economic Environment

2.1 Conceptual Framework

We consider a labor market facing two-sided limited information: firms and/or workseekers may imperfectly observe workseekers' skills. We show that these information frictions can distort job search decisions and job offer decisions, in turn distorting employment and earnings. The framework also predicts lower firm-level productivity and revenue but we do not focus on these implications, because we do not observe firm-level variation in information. We first describe the full-information labor market and then introduce information frictions on the supply and demand sides. We begin with a very simple static model and discuss extensions at the end of this section.

On the supply side, each workseeker i has one heterogeneous attribute, skill S_i, and homogeneous time endowment, T. Based on these attributes, she makes two sequential decisions. Her first decision is how to allocate time between job search W_i and other activities O_i. Job search yields utility $U^W(S_i; W_i)$, discussed in more detail below, while time spent on other activities yields utility $U^O(S_i; O_i)$. These utility functions include pecuniary and non-pecuniary benefits and costs such as wages, monetary costs of job search, and leisure.[7] She then receives a job offer with probability $p(S_i; W_i)$ with an associated wage $Y(S_i; W_i)$. We use 'wage' as shorthand to reflect both pecuniary and non-pecuniary aspects of the job. Her second decision is to accept or reject the job offer.

In a full information world, she chooses W_i and $O_i = T - W_i$ to equate the marginal return from job search and the marginal return from other activities

$$\frac{\partial U^W(S_i; W_i(S_i))}{\partial W_i} = \frac{\partial U^O(S_i; T_i - W_i(S_i))}{\partial W_i}. \tag{1}$$

If she receives a job offer, she accepts any offer with wage $Y(S_i; W_i) > Y(S_i; W_i)$. Her decision decisions may depend on her heterogeneous skill S_i in two ways. First, $W_i(\cdot)$ will be a non-degenerate function of S_i if skill and time are non-separable in the utility functions $U^W(\cdot;\cdot)$ and/or $U^O(\cdot;\cdot)$. For example, the return to job search may higher for workseekers with higher skills, in which case the optimal time allocation to searching for formal work will be increasing in skill. Second, her wage and/or reservation wage will depend on skill if firms' pay a skill premium (discussed below) or workseekers with higher skills have different outside options.

On the demand side, each firm f has a single fixed attribute, productivity A_f. The firm produces output Q_f using effective labor L_f^E and other inputs K_f with a constant elasticity of substitution (CES) production function that depends on productivity A_f. Firms face fixed output prices and costs of capital. Effective labor $L^E = \sum_{i \in f} S_i$ is the sum of the skills of all workers employed by the firm. Absent information frictions, firms' optimal choice of labor satisfies $p \frac{\partial Q(L^E;K)}{\partial L} = w$, where w is the wage. Workers are paid their marginal revenue product, so workers with higher skills receive higher wages. This model gives rise to a threshold rule, where each firm hires workers in a skill band that depends on the firm-specific technology A_f. In the full-information equilibrium, workseekers' time allocation and firms' input choices generate a wage for each skill level $Y(S_i; W_i(S_i))$ in which all workseekers who receive job offers accept them.

We now introduce supply-side information frictions. Workseekers observe a noisy proxy $\tilde{S}_i = F(S_i; \epsilon_i)$, where ϵ_i captures the information friction. Workseekers allocate time between job search and other activities based on $\tilde{S}_i$, but their payoffs from job search depend on S_i. Hence, the new indifference condition is

$$E\left[\frac{\partial U^W(S_i; W_i(\tilde{S}_i))}{\partial W_i}\right] = E\left[\frac{\partial U^O(S_i; T_i - W_i(\tilde{S}_i))}{\partial W_i}\right] \tag{2}$$

[7] The utility function $U^W(\cdot;\cdot)$ can be interpreted as the reduced-form of three structural functions: a job production function that maps skills and search time into a probability of securing a job, a wage function that maps skills into earnings conditional on securing a job, and a utility function over wage-financed consumption and time allocated to other activities. The wage function will reflect the dynamic process generating turnover and earnings trajectories. We focus on the reduced-form predictions for the hiring margin, as we do not observe longer-term labor market trajectories

where the expectation is taken over the distribution of ϵ_i. The optimal time allocations in conditions (1) and (2) will be equal if both utility functions are linear in skill and time. Otherwise, the time allocations will generally be different. If, for example, utility from job search is a concave function of time allocated to job search, then information frictions will lower the optimal time allocation to job search. This will in turn lower both the employment rate and earnings, provided these are increasing functions of job search time.

Demand-side information frictions occur when firms cannot perfectly observe workseekers' skills. The concavity of the CES production means that the expected marginal revenue product from each work seeker is lower than their expected skills:

$$E\left[p\,\frac{\partial Q\left(L^E;K\right)}{\partial L^E}\right] < p\,\frac{\partial Q\left(E\left[L^E\right];K\right)}{\partial L^E}. \tag{3}$$

If wages are completely flexible and all firms face information frictions, then wages will fall at each level of skills. If there is some floor on wages then the lower wages may not fully offset the fall in expected revenue and firms will reduce total labor demand.[8] Labor demand will also fall if firms face separation costs that drive up the cost of bad hires, such as firing regulations, mandatory severance pay, or fixed costs of recruiting and training new staff. Concavity of the production function is sufficient but not necessary to generate this result. The same result holds if firms are risk averse and cannot fully insure against negative shocks from hiring decisions, such as workers damaging equipment, alienating customers, or missing work when absenteeism is costly.

Demand-side information frictions can reduce wages even if wage floors, separation costs, or uninsurable risks are not sufficient to reduce employment. Information frictions by themselves reduce the expected and risk-adjusted marginal revenue products from new hires and firms have an incentive to pass this on to workers through lower wages.

It is possible that two-sided information frictions can interact to further distort labor market outcomes. Consider a simple example where workseekers and firms each observe independent signals of workseeker signals, respectively $\tilde{S}_i = S_i + \epsilon_i$ and $\hat{S}_i = S_i + \eta_i$. Each side knows that the other side faces information frictions but not what signal the other side sees. Workseekers' search decisions will be distorted because they do not know their own skills and because they know firms cannot assess their skills, increasing the risk of `incorrect' job and wage offers. Firms' offer decisions will be distorted because they do not know workseekers' skills and because they know workseekers cannot assess their own skills, increasing the risk of `incorrect' applications. Under certain parameterizations of the model, these two-sided frictions can distort employment by more than the sum of the two individual frictions.

[8] This wage floor may reflect minimum wages or positive reservation wages from workseekers' outside options. Even a wage floor at zero may distort hiring decisions if information frictions are very large and there is a positive probability that some workseekers will have negative marginal revenue products. Pallais (2014) also argues, using a different model structure, that information frictions can generate unemployment when there is a wage floor.

This simple framework illustrates that information frictions on either the demand or supply side of the labor market can generate lower employment, lower wages for the employed, and hence and lower earnings total. The framework also illustrates an important distinction. Eliminating demand-side information frictions can change labor market outcomes without any change in workseekers' behavior. Eliminating supply-side information frictions can change labor market outcomes only by changing workseekers' behavior. This motivates our measurement and analysis of workseekers' beliefs and job search in the skill certification experiment. However, skill certification can also change job search behavior if workseekers *believe*, correctly or incorrectly, that certification will change firm behavior. This motivates our second and third experiments that manipulate information available to respectively only the supply side and only the demand side of the market.

The framework does not generate a simple prediction about which types of workseekers will be affected most by information frictions. The incidence of distortions depends on the joint distribution of true skills and perceived skills and on the parameterization of the model. For example, if firms observe skills with classical measurement error and shrink observed skills toward the population mean, then information frictions will attenuate the return to skill. Alleviating frictions will raise high-skilled workseekers' wages and potentially employment and will hurt low-skilled workseekers' wages and potentially employment. Alternatively, the wedge between true and observed skills may be larger at low levels of true skills, perhaps because high-skilled workseekers can acquire other signals. In this case alleviating frictions will have limited effects on high-skilled workseekers. We view the incidence of distortions as an empirical question and test for heterogeneous effects of improved information provision on workseekers with different skills.

This framework can be easily generalized to relax several simplifying assumptions. We focus on single-dimensional skills. But the core results are unchanged if skills are multidimensional and at least one dimension is imperfectly observed. We focus on a single measure of search effort time and abstract away from multiple search strategies. But the core results are unchanged if workseekers can search in multiple sectors or use multiple strategies and information frictions will then also distort time allocations between different types of search. The core results are also unchanged if we allow pecuniary costs of search and perfect credit markets. With pecuniary credit constraints, distortions due to information frictions are more difficult to characterize and depend on the joint distribution of true skills, noise, and access to credit. We assume firms have no market power in the labor market. Relaxing this assumption does not change the prediction that information frictions distort search and hiring. Perhaps most importantly, we focus on a static framework that does not allow firms or workseekers to learn. This framework can be extended to a multiperiod model where firms observe workers' skills after hiring them, as in Altonji and Pierret (2001). As long as revelation is not instantaneous or firms incur hiring or firing costs, the predictions of the framework are unchanged. This simplification is motivated by the fact that learning cannot occur for unemployed workers and empirical work in the US shows that firms learn slowly about employed workers' skills (Arcidiaono et al., 2010; Lange, 2007).

2.2 Context

We work in the metropolitan area of Johannesburg, South Africa's commercial and industrial hub. Johannesburg's labor market has four salient features for our study, although none of these are unique to this setting. First, unemployment is common. Second, school qualifications are weak signals of skill, which might lead to firms or workers having limited information about workseekers' skills. Third, firms face high separation costs. Fourth, there is a binding minimum wage. Our framework suggests that the second feature, combined with either the third or fourth feature, can lead to lower employment and earnings. These features are common in the Middle East, North Africa, Latin America and Southern Europe.

First, employment is low, particularly for youths. In our study period unemployment was 28% for the working-age population, 51% for people aged 15-24, and 32% for people aged 25-34.[9] Most employment was in the formal sector, where at least some job search and hiring is through formal channels where skill certificates might be used.[10]

Second, South African firms report they find it difficult to screen workers or struggle to find workers with suitable skills.[11] Grades and grade progression in schools serving poor communities are weakly correlated with independently measured skills (Lam et al., 2011; Taylor et al., 2011; Van der Berg and Shepherd, 2015). This limits the signal employers obtain about skills from from grade attainment. There is only one nationally standardized assessment in South African education, a secondary school graduation examination. Workseekers typically report their grades on this examination in job applications. But qualitative interviews with 20 firms raised concern that grades in this examination convey little information about skills. Examination grades are also weakly correlated with performance in post-secondary education (Schöer et al., 2010). Perhaps because education signals are noisy, referrals from employees' social networks are widely used.[12] Certification is thus likely to provide firms with additional information on workseekers' skills, even conditional on educational attainment. Certification is also likely to give information to workseekers,

[9] Throughout the paper, we use Statistics South Africa's definition of an employed person as someone who did any income-generating activity, for at least one hour, during the reference week. Employment rates exclude those in education or not in the labor force.

[10] At the time our interventions were implemented, 77.39% of the employed in Johannesburg were in formal jobs, 15.6% in informal jobs and the remainder in agriculture or working for private households (Statistics South Africa, 2016). Informal jobs are defined as those with no written contract. The largest employers were finance (22% of workers), community and social services (21%), trade (20%) and manufacturing (13%).

[11] For example, in a survey of formal and informal SMEs, only 14% said workers have the skills demanded by business; 21% stated that workers had minor skills deficits and 52% said there were significant skills deficits (ILO, 2016). 16% of a sample of urban small and medium enterprises say finding people with suitable skills is the top factor inhibiting employment (Small Business Project, 2013). And a survey of the 100 largest firms lists availability of a skilled workforce as their top priority when deciding where to locate operations (World Economic Forum, 2018).

[12] 81% of employees at urban firms of less than 20 employees and 41% of employees at firms with more than 100 employees were recruited through referrals from existing employees (Schöer et al., 2014).

who may have received unreliable feedback on their performance in school. This weak correlation between educational attainment and skills applies in many other developing countries (Pritchett, 2013; Söderbom and Teal, 2004).

Third, many firms feel they face considerable risk from poor hires, partly due to high separation costs. Hiring procedures, such as restrictions on part-time or temporary contracts, and firing procedures are quite rigid (Botero et al., 2004; Bhorat and Cheadle, 20109). Separation is difficult, with stringent requirements to ensure dismissal is procedural. Employees, even those hired temporarily, can challenge dismissal in institutions dedicated to resolving labour disputes. Firms with less than 50 employees had had an average of two cases in the last year in dispute resolution, spending an average 11 days of staff time per case, which is costly even if they win the case (Rankin et al., 2012). Firms report being concerned and confused about labor regulation[13]. Indeed, firms offered free access to a newsletter and website about labor legislation increased hiring, highlighting that perceptions about regulation may reduce employment (Bertrand and Crépon, 2019).

Fourth, there is a wage floor and workers have outside options for income. There was no national minimum wage at the time of the study, but minimum wages were set by sector (ILO, 2016). These wages are not perfectly enforced. But compliance is higher in the formal sector (where 31% of workers earned below minimum wage in survey data, compared to 59% in the informal sector) and in dominant sectors in Johannesburg (Bhorat et al., 2016). An extensive system of social grants (mainly old-age pensions and child support grants) means many workseekers have some outside income (ILO, 2016).

There is thus some evidence that both firms and workseekers might have limited information about workseeker skills. Furthermore, there are high separation costs, a binding minimum wage or workers with reservation wages in this context. The model predicts these may result in lower employment and earnings if information frictions are present. We now test for the existence of these frictions experimentally.

2.3 Sample Recruitment and Data Collection

We recruit a sample of young, actively searching, unemployed and underemployed workseekers from low-income backgrounds with at most 12 months of cumulative work experience. They have limited access to traditional signals of productivity: university education, references from prior employment, or family connections. They are likely to have limited information about their skills. We deliberately focus on a theory-relevant population rather than seeking a representative sample[14].

[13] Only 18% of a random sample of firms with 10-300 workers knew the conditions that made a contract valid or how many months of pay were due to workers who were unfairly dismissed (Bertrand and Crépon, 2019). 54% of a sample of SME owners and 25% of a sample of informal enterprise owners stated that labor legislation is a major constraint on business growth (ILO, 2016). 15% of a separate sample of urban SME owners say labor regulations are the top factor inhibiting employment (Small Business Project, 2013).

[14] Our sample are active job searchers, which does not consider the impact of information frictions on the discouragement margin.

To recruit the sample, we work with the Harambee Youth Employment Accelerator, a social enterprise that connects large and medium-sized firms with first-time work seekers from low-income backgrounds. Harambee recruits candidates through radio and social media advertising and door-to-door recruitment in low-income neighborhoods. Interested candidates register with Harambee online. Harambee conducts a screening over the phone to assess if candidates meet their eligibility requirements and tells candidates the information may be checked against administrative data.[15] Eligible candidates are invited to two days of standardized assessments in downtown Johannesburg to evaluate their cognitive skills, literacy, numeracy, and aptitude for different career types. Roughly 2% of individuals with the top assessment results are invited to join training programs that place them in jobs with Harambee's partner employers, over 450 of South Afrca's larger firms.[16] Our sample consists of all candidates who arrive at Harambee for the second of these two testing days, on 84 operational days.[17]

We conduct two survey rounds to measure workseekers' labor market outcomes, search, and beliefs about their skills and the labor market. The baseline, a self-administered but supervised questionnaire on desktop computers, is held at Harambee. This is administered after candidates have sat the skills assessments but before they receive any information about their results. We collect endline data in a 25 minute phone survey administered by JPAL-Africa enumerators roughly 3-4 months after treatment.[18] The phone survey response rate is 96%, balanced across treatment groups (Appendix Table A.6), and related to few baseline covariates (Appendix Table A.7). We also collect two measures of workseekers' beliefs about their skills 2-3 days after treatment using a text message survey. Respondents receive payments, via mobile phone airtime, for answering both text message and phone surveys.

We report summary statistics for key baseline and endline variables for the 6,891 workseekers in our sample in Tables A.2 and A.3. Our target criteria are largely adhered to. Respondents are 24 years old on average, with the 90th percentile at 28 years. Only 1% had not completed high school. 17% of the total sample also have a diploma or degree and 21% have some type of post-school certificate.At baseline, only 9% had been paid a salary for long-term work.

[15] Consistent with our targeting criteria, Harambee requires that candidates be aged 18-29, have a school-leaving certificate, have not got more than 12 months of formal work experience and come from low-income homes. Harambee imposes the additional criteria that candidates must be South African citizens and have neither a criminal record not a credit score blacklisting.

[16] In our endline survey of 6,891 workseekers 3-4 months after they completed assessments with Harambee, only 1.39% had done further tests or interviews with Harambee, 0.61% had been selected for a Harambee training programme and 0.17% had received a job offer through Harambee.

[17] Harambee has a separate assessment stream for candidates with physical and certain learning disabilities, which is tailored to their needs (see Appendix A). These candidates are not included in our sample.

[18] See Garlick et al. (2019) for an experimental validation of labor market data from phone surveys in this setting.

2.4 Job Search and Employment in Our Sample

This section describes relevant patterns around labor market outcomes and job search in our sample. Only 38% of our sample are men. Women have lower employment rates and face more di cult transitions into the labour market, despite having higher levels of education, so may be more likely to apply to Harambee.[19]

Most workseekers have some work experience. Only 9% of respondents had held a formal job at baseline. But 55% had done short term casual or contract work, 26% had been self-employed, 13% had worked in an apprenticeship or internship and 21% had helped unpaid in a business. Only 20% had none of these experiences. A large portion of work is temporary, leading to high levels of job insecurity and many changes of job. 38% of the sample had worked in the past seven days at baseline. By endline, 16% of workseekers had not been working at baseline but had worked in the past seven days. 22% of workseekers had been in work at baseline but were not working at endline, 46% of the sample were not in work in either round. 16% were working in both rounds. Of the 16%, only 44% (i.e. 10% of the total sample) were in the same job. Jobs were often short in duration: of the employed at endline, tenure in their current job was a median of 1.93 months (mean 7.38).[20] Within wage work, hours are less than full-time. At endline, those working worked an average of 29 hours a week. Only 41% of those working worked a full time load (35 hours) or more. In focus groups, many workseekers reported being on shift or zero-hour contracts where they would like to work more. Similar trends hold in many developing countries (Donovan et al., 2018).

Conditional on working, respondents earn similar amounts to the minimum wage for low-skilled work in urban areas, an average of 560 ZAR (40 USD) per week at baseline.[21] Baseline earnings are roughly 1.82 times the most widely used weekly adult poverty line and 92% of the weekly minimum wage in urban areas for hospitality (or 78% of the wage for wholesale and retail) (see Appendix D.2).

Job quality is higher in wage jobs than in work for the family or self-employment, as is true in other developing countries (Donovan et al., 2018). At endline, 56% of those working were in wage jobs. They have higher earnings, of 889 ZAR in the past week on average, compared to 403 ZAR for those in family- or self-employment. 71% of those in wage work had written contracts, compared to 27% in family- or self-employment. Most wage employees worked for somewhat larger rms.[22] Among those in family rms or self-employment, 87% were in rms with four or fewer

[19] In a two year panel of workseekers aged 20-24 in three urban areas, one (two) years after baseline, men were 3.7 (7.4) percentage points more likely to be participating in the labour force, 10.7 (11) percentage points more likely to be in wage employment and earned R782 (R461) per month more than women (conditional on education and some measures of household earnings) (Levinsohn et al., 2013). Nationally, similar patterns hold and a higher proportion of women are in no or low skilled jobs (Statistics South Africa, 2013).

[20] A few longer term jobs did exist: for the 10% of workers in the same job at baseline and endline, tenure was a median of 9.86 months (mean 19.62).

[21] One ZAR is 0.16 USD PPP at the time of the baseline.

[22] The breakdowns of formal workers by rm size are: less than 4 employees - 27%; 5-19 employees - 10%, 20-49 employees - 22.67%, 50-200 employees - 14%, more than 200 employees - 27%.

employees, Accordingly, permanent work is more desirable: 49% of the sample were only searching for permanent jobs, while only 8% were only applying for temporary jobs. However, such jobs are hard to attain: only 15% of those over 28 had had a permanent job, although 66% had held short term work. At endline, only 3.5% of the employed had a permanent job with a written contract.

Most respondents are actively searching, with 97% reporting any job search activity in the previous 7 days at baseline. At baseline, respondents searched 17 hours in the past week and had submitted 9 job applications in past 30 days. They spent about 189 ZAR on job search in the past seven days, including transport and printing costs. Success rates are fairly low: workseekers received on average 0.74 responses and 0.822 offers in the previous 30 days. Given the high churn in and out of work, at endline search behavior is similar to baseline. In part, some candidates have moved out of work. But candidates also continue searching while working.

Finally, parts of the application process are somewhat formal. Those employed in the control group at endline were asked about how they obtained their last job. 41% got work through a process involving submission of a written application without a referral.[23] 48% got work through referrals: 35% said a social contact gave them the job directly, and a further 13% had an interview after a social contact referred them. But even workseekers applying through referrals may get their contact to pass on a written application and hence benefit from providing a certificate of skills.[24] Thus certificates to attach to a written application are a useful way of communicating information about skills to employers.

2.5 Skill Measures

We conduct six skills assessments with workseekers: numeracy, communication, concept formation, grit, focus, and planning. Appendix A describes what tests measure and their relationship with labor market outcomes and productivity. The numeracy, communication and concept formation assessments are used by Harambee to select which candidates they place in firms. We chose the grit, focus and planning measures after conducting focus groups with 20 HR managers in large and small firms. They mentioned soft skills of these types as valuable. We use tasks rather than self-reported scales because firms thought scales might be easy for respondents to game.

Firms seem to value information about workseekers' skills on these tests. First, they already use them. Harambee has used different combinations of the cognitive tests to select over 20,000 candidates for entry-level jobs since 2011. In Appendix Table A.1, we describe current use of these assessments by 33 large firms. All firms use at least one cognitive test to screen all their entry-level

[23] 20% dropped off a CV and then had an interview, 6% dropped off CV and got a job without interview, 6% emailed a CV or applied for a job online and 9% got a job through an employment agency or labor broker.

[24] We show later see that receiving a certificate on increased the probability of getting a job through a referral. More informal methods of getting jobs that do not require any proof of qualification are less common: only 8% of workseekers were given their job directly at the work site and 2% got a job after an interview without dropping off a CV.

candidates: 24 used all three tests, 2 used two and 7 used one. In contrast, only 20 (60%) required a CV and 19 (57%) required a matric certificate with test scores. Second, we show in Section 4.4 that other firms have positive willingness-to-pay for better information on workseeker skills in two small incentivized choice experiments. Firms value the cognitive and soft skills measures similarly. This suggests firms find this skill information as, if not more informative, than other traditional signals of skill for selecting candidates.

Given these patterns, why has a market-based solution not arisen to provide certificates of psychometric assessments? As in many markets, the fixed costs of in-house development may be too high, particularly for small firms (Bartram, 2004). These problems are particularly serious in countries with stringent regulation of psychometric testing, like South Africa and the US, where psychometric testing was used historically to facilitate racial discrimination (Foxcroft, 2004). In South Africa, legislation requires that employers must psychometrically validate any new test and only psychologists can administer tests (Foxcroft, 1997; Paterson and Uys, 2005). Entry by third party providers can also be difficult because they need to establish credibility.

In Appendix C, we propose a nonparametric test to establish if our multidimensional skill measures are related to candidates' demographic characteristics, job search, and labor market outcomes. We find that skills are higher for younger candidates, men, and candidates with post-secondary education, particularly university degrees. Skills are also positively correlated with self-esteem. Candidates with higher skills search more actively but they are not more likely to be employed. These patterns might reflective more selective search by higher-skilled candidates or an existing information friction that limits employers' scope to observe their skills.

2.6 Skills Assessment Process

Assessments are conducted over two days. Each assessment session is led by two to three psychologists, who manage a team of facilitators. Assessments are conducted in English and are self-administered on desktop computers.[25] The assessments yield six cardinal skills measures. Appendix Table A.2 shows standardized scores on tests. There is a fairly even spread of candidates over the distribution and little evidence of ceiling effects.

On the certificates, we display only the tercile in which a candidate placed on each assessment relative to a benchmark group of candidates.[26] There was no way for candidates to access their underlying score or for control group workseekers to access their scores.

Very few candidates score well or poorly on all tests, because skills are not highly correlated

[25] Before assessments, candidates do practice computer exercises. However, some candidates have poor computer skills, so assessment results will be driven in part by candidates' computer skills. The assessments have longer time limits to allow for poorer computer skills and no tests require fast responses. Before starting assessments, candidates consent to information on their skills being held by Harambee and shared with firms.

[26] We used data on 5,000 workseekers assessed at Harambee in Johannesburg before the study as a benchmark for the literacy, numeracy and concept formation test. For the remaining tests, we use data on 500 workseekers we assessed as a benchmark.

across tests. Only 0.7% of candidates have six bottom terciles and 2.3% have six top terciles. Most candidates have at least one bottom tercile (75.7%) and at least one top tercile (88.1%). The certificates thus allow candidates to differentiate horizontally across skills as well as vertically within skills.

Respondents have inaccurate information about their own skills. We measure their beliefs about which tercile they will be in for each cognitive skill after they have taken the assessments but before the treatment group receives their certificates. On average, workseekers predicted their tercile correctly for only 39% of assessments: they overestimated their score on 50% of assessments and underestimated it on 11%.

3 Labor Market Effects of Skill Certification

3.1 Intervention

We design a skill certification intervention which gives candidates information about their skills assessment results and allows them to signal the results to prospective employers. Conceptually, this is similar to existing signals like education or work experience. Like existing labor market signals, supply-side actors have information and endogenously choose whether to share this with the demand side. Like some but not all existing labor market signals, the demand side is unlikely to infer anything about the skills of candidates who do not share certificates. We work with a small fraction of the labor market and, to the best of our knowledge, there is no other provider of psychometric skill certifications in South Africa. We return to questions about spillovers in Section 4.3.

Candidates receive a package of 20 color copies of a certificate describing the assessments and their performance, printed on high-quality paper, an email with the certificate, and a group briefing with a psychologist (Figure 1). The certificate explains that Harambee has and placed candidates with over 250 firms in retail, hospitality, financial services, and other sectors. It notes that assessments are designed by psychologists and predict candidates' productivity and success in the workplace. It describes the six skill assessments and directs the reader to https://www.assessmentreport.info/ for more information on Harambee and the assessments. The certificate then shows the tercile in which the workseeker ranked on each assessment, compared to a benchmark group of other candidates assessed by Harambee.[27] To link candidates with certificates, each certificate shows the candidate's name and unique national identity number.[28] To provide credibility to the assessments and results, the certificate is branded with the World Bank and the Harambee Youth Employment

[27] In piloting with workseekers and firms, we found absolute scores provided readers with less information, as they could not easily anchor absolute scores to real outcomes. The benchmark group are described as young (age 18-34) South Africans assessed by Harambee who have completed secondary school and are from socially disadvantaged backgrounds.

[28] Candidates usually provide their identity documents, which contain their name and identity number, with job applications.

Accelerator logos.[29]

In the briefing, the psychologist explains how to interpret each of the skill measures on the certificate. They also explain that workseekers have the option of attaching the certificate to future job applications and can request more certificates from Harambee. A briefing script and Powerpoint presentation was jointly developed by the research team and the psychologists employed by Harambee. Research assistants monitored each briefing to ensure psychologists used the script.

All candidates, whether or not they receive the intervention, are told that very few candidates will be placed in jobs by Harambee and encouraged to continue searching for work. Candidates from all groups receive a limited workseeker support package: information on how to prepare and dress for an interview, how to create an email address, a CV template they can populate, and a list of job search strategies.

3.2 Experimental Design

We randomly divide workseekers into a skill certification and a control group. We randomize treatment by day to reduce risks of spillovers between treated and control workseekers.[30] Treated workseekers receive the certification intervention. Control group workseekers receive no detailed information, printed or emailed, on their performance on the individual tests. At the end of the second day of assessments, they receive an end-of-day briefing where they are told whether they will be invited to further stages of assessment with Harambee. This is the only information they receive on their assessment results.

We estimate treatment effects using models of the form

$$Y_{id} = T_d \quad + X_{id} \quad + \quad S_d + \quad {}_{id}; \tag{4}$$

where Y_{id} is the outcome for workseeker i assessed on date d, T_d is the treatment assignment, X_{id} is a vector of prespecified baseline covariates, and S_d is a stratification block fixed effect. We use heteroskedasticity-robust standard errors clustered by assessment date, the unit of treatment assignment. We apply an inverse hyperbolic sine transformation to right-skewed variables such as earnings; the distributions of these variables in our sample allow us to roughly interpret these treatment effects as percentage changes. We assign zeros to job characteristics for non-working respondents (e.g. earnings, hours) and to search measures for non-searching respondents (e.g. number of applications submitted) to avoid sample selection. All labor market and job search measures use 7-day recall periods, except where we specify otherwise.

We report treatment effects on prespecified outcomes in tables 1, 4, 5, A.8, and A.9 and discuss

[29] Harambee is a widely recognized brand in South African marketing surveys (Mackay, 2014).

[30] Randomization is sequential and stratified, with days randomized within blocks of 6-10 upcoming days. Table A.2 shows that the randomization generates balanced treatment assignments. There are respectively 2,247 and 2,274 workseekers in the skill certification and control group, spread over 54 days. The treatment occurs on the second day of assessments when workseekers have completed all tests.

Figure 1: Sample Public Certificate

Notes: This figure shows an example of the certificates given to candidates in the skill certification treatment. The certificates contain the assessment results, the candidate's name and national identity number, and the logo of the World Bank and the implementing agency. Each work seeker received 20 of these certificates and guidelines on how to request more certificates.

treatment effects on some additional outcomes in the text below. We account for multiple testing across outcomes in two ways. We group outcomes into prespecified families that are measures of similar concepts and should not be viewed as independent tests: employment, job attributes, certificate use, search effort, search effectiveness, and beliefs about skills. First, we report q-values that control the false discovery rate across outcomes within each family (Benjamini et al., 2006). Second, we estimate treatment effects on inverse covariance-weighted averages of the outcomes within each family (Anderson, 2008). This provides a single summary test of the information contained in each family.

3.3 Skill Certification Improves Average Labor Market Outcomes

Skill certification substantially increases employment and multiple measures of job `quality'. Current employment rises by 5.2 percentage points from a control group mean of 30.1 percentage points (Table 1 panel A column 1). This increase begins in the first month after treatment.[31] Total hours worked increase by 20% and weekly earnings by 34%; hence hourly wages increase by 20%. These are percentage increases relative to the control group mean, which includes zeros assigned to all non-workers. We benchmark effects on earnings in Appendix D.2. The treatment effects on earnings are large: they are roughly 17% of the weekly adult poverty line and 9% (7%) of the weekly minimum wage in urban areas for hospitality (wholesale and retail).[32]

Skill certification increases mainly wage employment (2.3 percentage points), rather than self or family employment. It marginally increases the rate of written contracts by 2 percentage points, an important measure of job formality in South Africa. We see no shift in other potential indicators of match quality, such as the probability that candidates want to stay in their current job or the length of tenure. Skill certification has no effect on the rate of written permanent contracts, but these are rare in our sample, so these may be difficult margins to move.

These results are robust to accounting for multiple testing. All treatment effect estimates remain significant at conventional levels when we use q-values rather than conventional p-values. Skill certification increases the employment and employment quality indices by respectively 0.14 and 0.11 standard deviations. None of the treatment effect estimates changes by an economically significant margin when we do not condition on the prespecified baseline covariates, though some standard errors are slightly larger.

[31] Current employment is measured as doing any work for cash or in-kind benefits in the preceding 7 days. Employment in each month relative to treatment uses the same definition but is asked about a full calendar month. This explains why the control group mean values for employment in the first and second months after treatment, 46 and 44%, are higher than the mean value for current employment, 31%. If candidates are working irregularly, the portion who had employment in any one week may be lower than the portion who had any employment in the last month. We do not see a downward trend in control group employment, as this is not systematically lower for candidates with longer lags from treatment to endline.

[32] Earnings conditional on working at endline in the control group are approximately 1.68 times the most widely used weekly adult poverty line and 85% (72%) of the weekly minimum wage in urban areas for hospitality (wholesale and retail).

Table 1: Treatment E ects on Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)
Panel A: Employment Status Measures					
	Employed	Month 1	Month 2	Hours o	Index
Treatment	0.052	0.036	0.058	0.201	0.138
	(0.011)	(0.011)	(0.014)	(0.052)	(0.025)
q: Treatment e ect = 0	0.001	0.001	0.001	0.001	
Mean outcome	0.309	0.464	0.437	8.852	-0.000
Mean outcome for employed				28.847	
# observations	6605	6602	6605	6596	6607
# clusters	84	84	84	84	84
Employment `Quality' Measures					
	Earnings o	Hourly wage o	Written contract	Index	
Treatment	0.338	0.197	0.020	0.106	
	(0.074)	(0.040)	(0.010)	(0.028)	
q: Treatment e ect = 0	0.001	0.001	0.020		
Mean outcome	159.364	9.844	0.120	-0.000	
Mean outcome for employed	518.291	32.283	0.392		
# observations	6587	6572	6573	6607	
# clusters	84	84	84	84	

Coe cients are from regressing each outcome on a vector of treatment assignments, randomization block xed e ects, and prespeci ed baseline covariates (measured skills, self-reported skills, education, age, gender, employment, discount rate, risk aversion). Heteroskedasticity-robust standard errors shown in parentheses, clustering by treatment date. Mean outcome is for the control group. All outcomes use a 7-day recall period unless marked with z (30-day recall period) or y (since treatment). Outcomes marked with o use the inverse hyperbolic sine transformation.

We can decompose the treatment e ects on job attributes into extensive margin e ects, explained by the rise in employment, and intensive margin e ects, explained by changes in the types of jobs. This distinction is important. If skill certi cation allows workseekers to make better search decisions or rms to make better o ers, then mean match quality should be slightly higher in the treatment group. We propose a simple but, to the best of our knowledge, novel decomposition. Intuitively, the extensive margin e ect on earnings is simply the average treatment e ect on employment multiplied by the mean earnings in the control group conditional on employment. The intensive margin e ect on earnings is the average treatment e ect on earnings minus the extensive margin e ect. The same argument applies to hours and contract type but is not meaningful for hourly wages, which are already an intensive-margin e ect. See Appendix B for a formal proof.

Job attributes, except for earnings, shift mainly at the extensive margin (Table 2). The hours and contract type e ects are explained entirely by the extensive margin. In contrast, the extensive and intensive margin e ects for earnings are respectively 27 and 7%, both signi cantly larger than zero. Skill certi cation both increases employment and allows workers to earn more conditional on employment without working longer hours. This is consistent with an improvement in match quality, if proxied by hourly wages, but not if proxied by contract status.

Table 2: Decomposition of Job Attributes into Extensive and Intensive Margins

	(1)	(2)	(3)
	Earnings o	Hours o	Written contract
Total e ect	0.338	0.201	0.020
	(0.073)	(0.052)	(0.010)
Extensive margin	0.269	0.188	0.020
	(0.059)	(0.041)	(0.004)
Intensive margin	0.069	0.013	-0.000
	(0.040)	(0.020)	(0.008)
Treatment e ect conditional on employment	0.195	0.036	-0.001
	(0.113)	(0.058)	(0.024)
Control group mean	5.177	3.624	0.392

This table reports decompositions of treatment e ects on job characteristics into extensive and intensive margins. The extensive margins are the treatment e ects on job characteristics due to the treatment e ect on employment, evaluated at the mean job characteristics for the control group. The intensive margins are the residual treatment e ects on job characteristics, which must be due to changes in job characteristics for the employed candidate in the treatment group. The conditional e ect is the implied mean change in job characteristics per employed treatment group candidate. Heteroskedasticity-robust standard errors are shown in parentheses, clustering by treatment date. All outcomes use a 7-day recall period. Outcomes marked with o use the inverse hyperbolic sine transformation.

3.4 Skill Certi cation Has Slightly Better Labor Market E ects for Lower-Skilled Workseekers with Worse Labor Market Prospects

Heterogeneous e ects of skill certi cation by skill level are likely, as we explain in Section 2.1. The treatment is inherently di erent for workseekers with di erent skill levels. The treatment may also be more relevant for workseekers facing larger uncertainty about their skills, such as workseekers with less education or prior work experience. The distribution of e ects is important for welfare and for understanding which workseekers are likely to take up certi cation when it is o ered.

We nd mixed evidence of heterogeneous treatment e ects on labor market outcomes by skill, but demonstrating this pattern is complicated. We observe multiple measures of skill. Studying treatment e ect heterogeneity using standard methods requires collapsing these measures into a single index (e.g. using principal components analysis) or estimating models with separate treatment e ect parameters for all di erent skill measures and their interactions. The former approach imposes strong parametric assumptions (e.g. no complementarity between skills) and the latter approach raises dimensionality problems.

We instead use a nonparametric test based on the idea of rst-order stochastic dominance of skill distributions. We estimate the treatment e ect within each possible cell de ned by a combination of skill tercile values (e.g. top tercile for all skills). For each cell, we then estimate the di erence between the treatment e ect in the cell and the average treatment e ect in all cells within strictly lower terciles on all skills. If treatment e ects are increasing in skill, then this di erence should

Table 3: Heterogeneity in Labor Market Treatment E ects by Skill

	(1) Employed	(2) Hours o	(3) Earnings o	(4) Hourly wage o	(5) Written contract
Treatment e ect relative to units with strictly					
lower skills	-0.005	-0.074	0.156	0.133	0.010
	(0.029)	(0.100)	(0.149)	(0.083)	(0.012)
higher skills	0.051	0.178	0.105	0.049	0.000
	(0.024)	(0.090)	(0.148)	(0.094)	(0.013)
p: Di erences equal	0.139	0.065	0.801	0.495	0.571

This table reports estimates of treatment e ect heterogeneity using dominance tests. To implement these tests, we estimate the average treatment e ect within each cell de ned by the Cartesian product of the terciles of all skills. We estimate the di erence between the cell-speci c treatment e ect and the treatment e ects in all cells with strictly lower terciles of all skills (rows 1-2). We estimate the di erence between the cell-speci c treatment e ect and the treatment e ects in all cells with strictly higher terciles of all skills (row 3-4). If the former quantity is larger than the latter, then treatment e ects are increasing by skill. Row 5 reports a p-value for testing equality of the two quantities. Heteroskedasticity-robust standard errors are shown in parentheses, clustering by treatment date and skill group. All outcomes use a 7-day recall period. Outcomes marked with o use the inverse hyperbolic sine transformation.

be positive. Similarly, we estimate the di erence between the treatment e ect in the cell and the average treatment e ect in all cells within strictly higher terciles on all skills. If treatment e ects are increasing in skill, then this di erence should be negative. See Appendix C for a more formal and comprehensive description of the test.[33]

Skill certi cation weakly attened labor market outcome-skill gradients (Table 3). Treatment e ects relative to cells with strictly higher skills are positive for employment and hours worked. This is consistent with negative heterogeneous treatment e ects by skill. The dominance test suggests a nonlinear relationship between skill certi cation e ects and job quality measures but these relationships are not statistically signi cant at conventional levels. In contrast to our nding, standard limited information models in labor economics predict that richer information environments should lead to steeper labor market outcome-skill gradients (Altonji and Pierret, 2001; Farber and Gibbons, 1996).

Skill certi cation e ects are also slightly smaller for candidates with better labor market prospects. E ects on employment and earnings are smaller, though not always statistically signi cantly smaller, for candidates employed at baseline, with post-secondary education, and with higher latent propen-

[33] Reporting skills by tercile on the certi cates identi es another type of heterogeneous treatment e ect. Candidates with skills near the cuto s between the top and middle terciles and between the middle and bottom terciles have very similar skills but receive di erent signals on their certi cates. We can use a regression discontinuity design with these candidates to analyze the treatment e ect of receiving di erent signals conditional on skill. Candidates near the top and bottom of the same tercile have di erent skills but receive the same signal on their certi cates. We can augment equation (4) with an indicator for having above-median skills in one's tercile, interacted with the treatment indicator, to analyze how treatment e ects vary across skills, conditional on signal. Provisional analysis does not nd systematically stronger evidence of heterogeneity using one approach ahead of the other.

sities for employment and high earnings.[34] These patterns show that skill certification is more helpful for candidates with worse labor market prospects, which may offset any latent tendency for skill certification to steepen the outcome-skill gradient.

These two forms of heterogeneity are not independent because skills, education, and baseline employment are related. We therefore run the dominance test separately for candidates with and without post-secondary education, with and without baseline employment, above- and below-median values of the latent propensity for employment, and above- and below-median values of the latent propensity for high earnings. Skill certification has a larger effect on employment for lower-skilled candidates within each subgroup, although the differences are not always statistically significant at conventional levels. The equivalent relationship for earnings differs across subgroups and differences are never statistically significant at conventional levels.

These patterns of heterogeneity suggest that skill certification does not steepen the outcome-skill relationship, even conditional on other proxies for labor market proxies. This might occur because candidates with lower skills are also less able to navigate job search or find other ways to signal their skills and thus benefit more from certification. This might also occur because some firms demand lower-skilled candidates, potentially because they are seen as good matches for some positions or firms are concerned that higher-skilled candidates are unattainable or will quit quickly.

Taken together, these results show that skill certification raises employment and raises earnings at both the extensive and intensive margins, with slightly larger effects for candidates with lower skills and worse labor market prospects. In the next section we explore what combination of supply- and demand-side responses generated these results.

4 Mechanisms Through Which Certification Changes Labor Market Outcomes

Skill certification can change labor market outcomes through four types of mechanisms. First, certification may solely alleviate distortions in firms' behavior. Candidates may have perfect information about their skills, while firms observe skills with error. Certain types of firm-side information frictions can lead firms to hire fewer workers and pay workers less, and alleviating these may match the treatment effects reported in Sections 3.3 and 3.4. Second, certification may solely reduce distortions in workseeker behavior. Firms may observe prospective workers' skills perfectly, while candidates observe their skills with error. Certain types of workseeker-side information frictions can distort candidates' job search decisions, leading them to receive fewer job offers and accept jobs that pay less, matching the estimated treatment effects. Third, there may be information frictions on both sides of the market, distorting both job search and hiring decisions. Fourth, skill certification may

[34] We estimate the latent propensities following Abadie et al. (2018). We regress each of employment and earnings on baseline demographics, education, skills, beliefs about skills, employment, earnings, and search behavior in the control group. We use the predicted values from these regressions in all treatment groups as latent propensities for employment and high earnings, adjusting the predicted values in the control group using leave-one-out estimation to avoid overfitting.

not provide useful information about candidates' skills to either side of the market but may make job applications more noticeable to firms. For example, firms might use a heuristic screening process for job applications that just selects colorful or professional-looking applications.

To test these mechanisms, we first examine treatment effects of skill certification on candidates' beliefs, job search activities, and job search outcomes. Changes on these margins strongly suggest but do not conclusively prove that workseekers have limited information about their skills. We then conduct additional experiments that separately vary information available to firms and to workseekers. We use these to identify the effects of one-sided information revelation and compare these to the effects of skill certification. Finally, we conduct a small experiment that tests the fourth, non-information-based mechanism.

These different mechanisms may have different implications for policy design, as outlined in the introduction. These different mechanisms may also have different implications for welfare, though we do not formally model the incidence of welfare costs of information frictions. For example, if information frictions only distort firms' choices, then information frictions can deliver at least temporary welfare gains to low-skilled candidates that firms incorrectly classify as high-skilled and pay accordingly. The net result may be a transfer from firms and high-skilled candidates to low-skilled candidates. If only workseekers face information frictions, then firms can pay high-skilled workers below their marginal revenue product. The net result may be a transfer from high-skilled candidates to firms.

The exact incidence of welfare costs and optimal policy design depend on the exact form of information frictions and any learning that takes place during job search and employment. We do not propose a formal quantitative model, which would need to impose assumptions about both candidates' and firms' objective functions. We merely argue that understanding which sides of the market face frictions is a necessary step for understanding the welfare implications of any frictions.

4.1 Skill Certification Changes Job Search and Beliefs

Skill certification can change job search behavior if it provides information to candidates (alleviating a real supply-side information friction). Alternatively, it can allow candidates to more credibly give information they already have to firms (alleviating a perceived demand-side friction). Both mechanisms may cause candidates to update their perceived return to specific job search investments. The two mechanisms are not mutually exclusive.

We first document that skill certification has a large effect on candidates' beliefs about their skills (Table A.8). We ask candidates if they think they were in the top, middle, or bottom third of candidates on each skill assessment and can compare this to their actual scores. We measure beliefs at baseline (after candidates have taken assessments but before they receive any information), by text message 2-3 days after treatment, and at endline. The outcome is a score out of 6 measuring on the number of skills for which a candidate correctly report the tercile they are actually in.

Certification increases the number of correct terciles from 2.3 to 3.3 out of 6. A similar update is visible in the text message survey conducted 2-3 days after treatment.[35] Skill certification also decreases how long candidates expect they will need to search before finding a job and increases the number of jobs they expect to receive. Treatment effects on self-esteem in the text message survey and the endline are both precisely estimated zeros, suggesting certification adjusted targeted beliefs about skills and employment prospects rather than general self-evaluation.

Candidates use skill certificates heavily (Table 4 panel A). 70% of candidates use the certificates with at least one job application, with an unconditional average of 6.7 applications sent per candidate. These applications generate an average of 0.43 interviews and 0.11 job offers over the 3-4 months from treatment to endline. All measures of certificate use are increasing in skill, using the same heterogeneity test described in Appendix C. Higher certificate use but slightly lower employment effects for high-skilled candidates is consistent with a higher per-application employment effect for low-skilled candidates. This is consistent with the lower levels of work experience and education in the lower-skilled group, which may limit their scope to signal their skills using other mechanisms.

However, treated candidates do not search more or more effectively in the week or month before the endline survey. Skill certification does not increase the probability of searching at the extensive margin, number of applications submitted, hours spent searching, money spent on search, or an inverse covariance-weighted average of these measures (Table 4 panel B). Similarly, skill certification does not increase the number of responses or offers received from employers in the preceding month (Table 4 panel C).

The combination of positive employment effects, high certificate use, and zero search effects appears to occur because search and employment rise soon after treatment and the endline search measures do not capture this. Employment rises by 3.6 percentage points in the first month after treatment and another 2.2 percentage points in the second month (Table 1 panel A). The questions on certificate use ask about the entire period between treatment and the endline survey, which covers the period when employment was rising. All other search measures ask about the preceding 7 or 30 days. This mostly misses the the first two months after treatment, as 78% of candidates completed the endline more than 90 days after treatment. Consistent with this timing hypothesis, effects on all search effort and search effectiveness measures are larger for respondents with a shorter time lag between treatment and endline. In particular, treatment increases the number of offers in the past 30 days by 0.09 (50% of the control group mean) for workseeker surveyed before the median treatment-to-endline time lag and has a tiny effect for workseekers surveyed after the median treatment-to-endline time lag. This result is robust to instrumenting baseline-to-endline lag with

[35] We do not report detailed results on belief updating, as this is not the primary focus of the paper. In brief: candidates who are overconfident at baseline update beliefs downward, candidates who are underconfident at baseline update their beliefs upward, overconfident candidates update less than underconfident candidates, and updating is not dramatically different across skill dimensions.

Table 4: Treatment E ects on Report Use

	(1)	(2)	(3)	(4)	
Panel A: Certi cate Use Measures					
	Any use y	Applications oy	Interviews y	O ers y	
Treatment	0.699	1.683	0.432	0.112	
	(0.013)	(0.040)	(0.023)	(0.011)	
q: Treatment e ect = 0	0.001	0.001	0.001	0.001	
Mean outcome	0.000	0.000	0.000	0.000	
# observations	6607	6596	6595	6595	
# clusters	84	84	84	84	
Panel B: Search E ort Measures					
	Any search	Apps yo	Search hours o	Search cost o	Index
Treatment	-0.020	0.018	-0.034	-0.092	-0.013
	(0.014)	(0.042)	(0.048)	(0.081)	(0.032)
q: Treatment e ect = 0	1.000	1.000	1.000	1.000	
Mean outcome	0.694	12.357	9.774	112.667	0.000
# observations	6606	6575	6599	6597	6606
# clusters	84	84	84	84	84
Panel C: Search E ectiveness Measures					
	Responses z	O ers z	Index		
Treatment	0.049	0.007	0.025		
	(0.047)	(0.017)	(0.029)		
q: Treatment e ect = 0	1.000	1.000			
Mean outcome	0.772	0.182	0.000		
# observations	6591	6590	6591		
# clusters	84	84	84		

Coe cients are from regressing each outcome on a vector of treatment assignments, randomization block xed e ects, and prespeci ed baseline covariates (measured skills, self-reported skills, education, age, gender, employment, discount rate, risk aversion). Heteroskedasticity-robust standard errors shown in parentheses, clustering by treatment date. Mean outcome is for the control group. All outcomes use a 7-day recall period unless marked with z (30-day recall period) or y (since treatment). Outcomes marked with o use the inverse hyperbolic sine transformation. We do not construct an inverse-covariance weighted average of the certi cate use measures because these do not vary in the control group, making it impossible to estimate the control group covariance matrix.

the random order in which candidates were assigned to be surveyed.

Treatment effects on on-the-job search are also consistent with this explanation. Skill certification appears to directly shift candidates from non-employed search into employment. Certification decreases the probability of non-employed search by 4.7 percentage points (standard error 1.5) and increases the probabilities of non-searching employment and searching while employed by respectively 2.5 and 2.7 percentage points (standard errors 0.9 and 1.0). There is no effect on the probability of simultaneously not searching and not working.

There are no certification effects on any of the search strategies we measure searching in a network, visiting businesses, and looking at advertisements. However, skill certification does slightly increase the probability of securing a job through a formal application or interview after a referral. There is no large or significant treatment effect on the probability of securing a job in other ways we measure: by approaching an employer in person, dropping off an application, emailing an application, getting hired by a social contact directly, or working at an employment broker. This suggests that the certificate might enhance the effectiveness of referrals, by encouraging network links to offer referrals or making their referrals more credible to employers.

Taken together, these results are consistent with candidates updating their beliefs about their skills, using their certification to assist their search soon after treatment, quickly increasing their employment rate, and then decreasing their search effort back to the control group level. This suggests that supply-side responses drive part of the certification effect on employment but that these responses will not necessarily persist.

4.2 Information Provision to Workseekers Has Some Labor Market and Search Effects

As supply-side responses appear to drive part of the employment effect of certification, are similar effects possible when information is revealed only to the supply side of the market? We test this with a `private' treatment arm of another 2,114 candidates in the skill certification experiment. These candidates receive one copy of an unbranded, anonymous certificate with the skill assessment results rather than multiple copies of a branded, identifiable certificate (Figure 2). They receive almost the same group briefing with the same industrial psychologists as candidates in the skill certification treatment. But the briefing only explains how to interpret the assessment result and how it might inform their own job search, without any encouragement to share the results with prospective employers.

We interpret the private treatment as primarily providing information to the candidates about their own skills. Candidates can share the private certificate with firms but this is less likely to change firms' decisions than the `public' skill certificates: the private certificates are not linked to a specific candidate (no name or national ID number), use Harambee's name but not any branding, and are not branded by the World Bank.

The private treatment changes candidates' beliefs about their own skills (Table A.8). These

Figure 2: Sample Private Certi cate

Notes: This gure shows an example of the certi cates given to candidates in the private treatment arm. The certi cates contain the candidate's assessment results but no identifying information and no branding. Each candidate received one copy of this certi cate effects are very slightly smaller than the corresponding skill certification effects on beliefs at the endline but identical at the text message survey. The private treatment marginally decreases expected search time and increases the number of job offers expected but by less than the public treatment. This is consistent with candidates perceiving the public certification as more useful for search.

Candidates in the private certification group do share their certificates with employers, but substantially less often than candidates in the public certification group (Table 5 panel B). Only 29% of candidates in the former group use the certificates in any applications (versus 70%) and they obtain on average 0.14 interviews (versus 0.43) and 0.04 job offers (versus 0.11) from these applications. Like the public certification, the private certification has no effect on job search behavior or outcomes in the week or month before the endline.

Private certification changes some labor market outcomes but by substantially less than public certification. Private certification has small and statistically insignificant effects on current employment and hours but slightly increases employment in the first month after treatment. It does more at the intensive margin, increasing weekly earnings by 16%, hourly wages by 10%, and the probability of having a written contract by 1.7 percentage points. All but the last effect are both significantly smaller than the corresponding public certification effects. We cannot reject that the intensive margin effects of the private and public certificates are equal, using the decomposition from Appendix B.

These findings are consistent with multiple possible interpretations. Private certification may convey the same information as public certification but the information may be less credible to firms. This would explain the similar effects of the two interventions on beliefs, the lower use of the private than public certificates in job search, and the smaller but sometimes still positive private effects on labor market outcomes. Alternatively, private certification may convey information only to workseekers and hence change beliefs about returns to search, job search, and labor market outcomes. As the private certification effects are mainly at the intensive margin, any changes in search behavior must generate higher-paying job offers, rather than more job offers. This might be driven by candidates applying to jobs where their skills are better matched. Future drafts will examine whether public and/or private certification increases the probability that candidates apply for jobs better suited to their skills.

We cannot rule out some firm-side learning in response to private certification. However, we argue that the public-private-control comparison is still useful. The three-way comparison shows that workseekers face information frictions and that giving workseekers better technology to signal their skills to firms improves their labor market outcomes. Giving workseekers information without any ability to share it with firms may be possible: we could give them information, measure their willingness to search and directed search decisions, and then send applications on their behalf to firms. But this does not correspond to any realistic policy intervention.

Table 5: Public and Private Certi cation E ects on Labor Market and Job Search Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Labor Market Outcomes												
	Employed	Month 1	Month 2	Hours o	Index	Earnings o	Hourly wage o	Written contract	Index			
Public	0.052	0.036	0.058	0.201	0.138	0.338	0.197	0.020	0.106			
	(0.011)	(0.011)	(0.014)	(0.052)	(0.025)	(0.074)	(0.040)	(0.010)	(0.028)			
Private	0.011	0.028	0.008	0.066	0.050	0.162	0.095	0.017	0.065			
	(0.012)	(0.013)	(0.015)	(0.048)	(0.028)	(0.078)	(0.046)	(0.009)	(0.030)			
q: public = 0	0.001	0.001	0.001	0.001		0.001	0.001	0.020				
q: private = 0	0.522	0.136	0.522	0.346		0.067	0.067	0.067				
q: pub = pvt	0.003	0.128	0.003	0.008		0.047	0.047	0.344				
# observations	6605	6602	6605	6596	6607	6587	6572	6573	6607			
# clusters	84	84	84	84	84	84	84	84	84			
Panel B: Job Search												
	Certi cate Use				Search E ort					Search E ectiveness		
	Any use y	Applications oy	Interviews y	O ers y	Any search	Applications yo	Hours o	Cost o	Index	Responses z	O ers z	Index
Public	0.699	1.683	0.432	0.112	-0.020	0.018	-0.034	-0.092	-0.013	0.049	0.007	0.025
	(0.013)	(0.040)	(0.023)	(0.011)	(0.014)	(0.042)	(0.048)	(0.081)	(0.032)	(0.047)	(0.017)	(0.029)
Private	0.290	0.572	0.144	0.036	-0.006	0.036	-0.035	-0.031	0.006	0.041	0.017	0.021
	(0.012)	(0.033)	(0.017)	(0.008)	(0.014)	(0.038)	(0.049)	(0.088)	(0.032)	(0.042)	(0.017)	(0.028)
q: public = 0	0.001	0.001	0.001	0.001	1.000	1.000	1.000	1.000		1.000	1.000	
q: private = 0	0.001	0.001	0.001	0.001	1.000	1.000	1.000	1.000		0.490	0.490	
q: pub = pvt	0.001	0.001	0.001	0.001	1.000	1.000	1.000	1.000		1.000	1.000	
# observations	6607	6596	6595	6595	6606	6575	6599	6597	6606	6591	6590	6591
# clusters	84	84	84	84	84	84	84	84	84	84	84	84

Coe cients are from regressing each outcome on a vector of treatment assignments, randomization block xed e ects, and prespeci ed baseline covariates (measured skills, self-reported skills, education, age, gender, employment, discount rate, risk aversion). Heteroskedasticity-robust standard errors shown in parentheses, clustering by treatment date. All outcomes use a 7-day recall period unless marked with z (30-day recall period) or y (since treatment). Outcomes marked with o use the inverse hyperbolic sine transformation. q-values control the false discovery rate within each family of outcomes.

4.3 Information Provision to Firms Has Some Effects on Responses to Applications

We conduct an audit-style experiment to identify the effects of providing more information about candidates' skills to only the demand side of the market. This experiment sends applications to real jobs on behalf of real candidates in our sample and randomizes whether applications include skill certifications. We vary the information available to firms when they review applications, directly replicating one stage in the job search process. This approach has one important advantage and three limitations. The advantage is that this isolates the firm's response to information and cannot be mediated by supply-side behavior (e.g. workseekers deciding which jobs to apply to and when to use certificates). The labor market results in public certification experiment might occur because workseekers think firms face information frictions and hence change their search behavior. The audit study allows us to test if firms actually face information frictions.

However, this does not provide a comprehensive test of all possible firm responses. The first limitation is that we observe only initial responses to job applications in the audit study, not job offers. The second limitation is that the audit study captures only undirected search. We do not systematically match candidates to jobs, which would happen to some extent in directed job search. Third, the audit study only captures firms' responses to one type of hiring strategy, which is not necessarily an important hiring channel for candidates in our sample. Only 6% of employed candidates in our skill certification experiment obtain jobs through online applications and treatment does not change this number. Treatment only substantially increases the probability of finding jobs in one way: formal applications after referrals, consistent with an important role for referrals. We therefore interpret the audit experiment as testing for the existence of information frictions facing firms hiring through one potential channel, but note that firms may face frictions when hiring through other channels too.

We describe the experiment briefly here, with more details in Appendix F. We invite a random sample of assessed candidates to send us a job application that we will forward to prospective employers on their behalf. We create a list of job vacancies by scraping and inspecting online job advertisements. We exclude advertisements that appear to be scams and restrict the remaining vacancies to jobs that might hire entry level workers without university education, such that all candidates in our sample would be eligible to apply. We do not attempt to match vacancies systematically to the attributes of candidates in our sample. We submit job applications from four randomly chosen candidates to each vacancy, each from a different email address and separated by a few hours. We manually code responses as `interview invitations,' `other response,' or no response. Automated responses are not classified as responses. `Other responses' are typically requests for more information from the candidate. We pass on all responses and interview invitations to the candidates. The current sample includes 3428 applications from 561 candidates sent to 857 vacancies but data collection for the experiment was still in progress in July 2019.

We generate between six and ten applications from each candidate and randomly assign half of these to be sent with certificates. Vacancies are randomized to receive either one or three applications with certificates identical to those used in the certification experiment. We randomly match applications to vacancies but never send multiple applications from the same candidate to a vacancy. The vacancy-level variation provides a partial test for crowd-out effects. This generates three layers of experimental variation: within-candidate variation in application-level treatment status, within-vacancy variation in application-level treatment status, and across-vacancy variation in the share of treated applications. This design naturally leads to the estimating equation

$$Y_{iv} = \text{Certificate}_{iv} \times \beta_1 + \text{Certificate}_{iv} \times \text{HighIntensity}_v \times \beta_2 + E_{iv} + V_v + \varepsilon_{iv}; \qquad (5)$$

where Certificate_{iv} is an indicator equal to one for applications submitted with certificates, HighIntensity_v is an indicator equal to one for vacancies that receive three applications with certificates, E_{iv} is a vector of email address fixed effects, and V_v is a vector of vacancy fixed effects. β_1 measures the effect of using a certificate when no other applicant does so, while β_3 measures the differential the effect of using a certificate when two other applicants also do so. The vacancy-level treatment assignment is omitted because it is colinear with the vacancy fixed effects. We use heteroskedasticity-robust standard errors clustered by vacancy, the highest level of treatment randomization, following Abadie et al. (2017). Results are similar when we cluster standard errors by candidate, condition on candidate-level fixed effects, or condition on candidate-level covariates.

We estimate treatment effects considering the full sample of vacancies (All), as well as dropping vacancies for which no application receives a response (Active), and additionally dropping vacancies for which all applications receive the same response (Selective). Estimates for the sample of active and non-automated vacancies are reported to address concerns that vacancies may have been filled in the interim or that responses are not legitimate.[36]

Applications with certificates are 1.4 percentage points more likely to receive responses and 0.9 percentage points more likely to receive interviews (Table 6). Both effect sizes are roughly 10% of the outcome mean but neither is significantly different to zero at conventional levels (p-values respectively 0.186 and 0.157). This treatment effect is partly driven by the 79% of vacancies that respond to no applications. These may arise because the firms review all our applications and choose not to reply. But it might also occur because vacancies are filled before we can collect applications from candidates and submit them. Few vacancies include explicit closing dates, so this hypothesis is difficult to test directly. When we restrict the sample to active vacancies that send a response to at least one application, the treatment effects rise to 8.2 and 5.2 percentage points. When we further restrict the sample to selective vacancies that do not send the same response to

[36] Neumark (2018) discusses the implications of employing the full sample of units (Bertrand and Mullainathan, 2004; Rich, 2014) or discarding matched units in which neither gets a response (Riach and Rich, 2002) for the interpretation of estimated effects.

all applications, the treatment effects rise to 12.5 and 6.5 percentage points.[37] None of these effects are significantly different to zero at conventional levels (p-values 0.168 - 0.248). We interpret this as evidence that information is marginally useful in the context of undirected search and may become more important when search is more directed. But the large standard errors mean this is a cautious conclusion.

Table 6: Treatment Effects of Additional Information in Audit Study

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Any response			Interview request		
Sample	All	Active	Selective	All	Active	Selective
Certificate	0.014	0.082	0.125	0.009	0.052	0.065
	(0.010)	(0.059)	(0.094)	(0.007)	(0.036)	(0.056)
Certificate HighIntensity	-0.021	-0.118	-0.185	-0.012	-0.065	-0.086
	(0.016)	(0.076)	(0.118)	(0.010)	(0.049)	(0.073)
Outcome mean	0.131	0.685	0.517	0.083	0.435	0.289
# applications	3428	713	466	3,428	713	466

Coefficients are from regressing each outcome on a vector of treatment assignments, vacancy fixed effects, and email address fixed effects. Heteroskedasticity-robust standard errors shown in parentheses, clustering by vacancy. The interview request outcome in columns 4-6 is a subset of the any response outcome in columns 1-3. The active sample in columns 2 and 5 exclude any vacancies where no applications receive responses. The selective sample in columns 3 and 6 excludes any vacancies where no applications or all applications receive responses.

The treatment effect of using a certificate decreases when competing against other applications with certificates ($\hat{}_3 < 0$). However, the mean response rate is still 2.6 percentage points higher in vacancies that receive three applications with certificates rather than one. This is consistent with a positive effect of better information on application-level response rates that is partly but not completely crowded out when applications are more widely used.

4.4 Firms Value Access to Information about Candidates' Skills

If better information about candidates' skill changes firms' decisions, then they should value this information. We test this claim using two firm-facing experiments. These are secondary experiments to test an additional prediction from our interpretation of our primary experiments. We therefore use relatively small samples and report the design and results briefly.

We recruit a sample of 69 firms by calling firms in an existing small business panel and by knocking on doors in a commercial areas near low-income residential areas in Johannesburg. We ask firms if they are willing to participate in a research study on hiring and tell them we can provide some useful information on hiring. We restrict the sample to business units that have hiring responsibilities (e.g. excluding branches of a larger firm that hires centrally). The firms in these

[37] This result follows mechanically from the underlying econometrics. The full-sample estimates of ${}_1$ in equation (5) are weighted averages of zero and the estimate in the sample of vacancies with any outcome variation (i.e. the sample in columns 3 and 6). The weights equal the shares of vacancies respectively with and without outcome variation. A similar argument applies to ${}_2$.

experiments may differ from the firms in our audit study and the firms where our workseekers apply for jobs.

We administer a survey about each firm's type, labor force, hiring practices, and hiring plans. We then conduct the two experiments. We frame both as research activities that are designed to help firms hire and that we hope to scale.

Our first experiment measures firms' willingness-to-pay for access to a secure online database containing assessment results and contact information for our candidates. This database allows firms to filter and search for candidates with specific skill profiles and obtain their contact information. See Figures A.2 and A.3 for selected screenshots. We measure willingness-to-pay using a variant of the Becker-DeGroot-Marschak mechanism. We tell firms the `normal' access price, ask how much they are willing to pay for access, then randomly offer them a discount between 0 and 100%, and give them access if their stated willingness-to-pay is higher than the normal price minus the discount. If their stated willingness-to-pay is below the normal price minus the discount, we give them access to a placebo database with candidates' contact information and selected resume-style information but no skill assessment results. We first explain the entire mechanism and run a practice round with a bar of chocolate.

Firms have a substantial willingness to pay. 70% of firms report a positive willingness to pay, with a conditional mean valuation of USD269. This is 54% of the mean monthly earnings for candidates in our workseeker sample. The product we offer is new and firms may be risk-averse, so these valuations are probably lower bounds.

Our second experiment measures firms' ordinal preferences for different types of information about candidates. We ask firms to rank applications from seven hypothetical candidates and tell them we will use their ranking to match them with candidates on the database, following Kessler et al. (2019). Applications randomly vary in their level of education (less than complete secondary school, complete secondary school, complete secondary school and a one-year diploma course) and skill profiles (all middle terciles, one high tercile, one low tercile).

Almost all firms value higher skills in any dimension above a post-secondary diploma. On average, firms value higher skills slightly less than completing secondary school (relative to the category `some secondary school education'). But there is substantial variation in firms' ranking of skills relative to each other and relative to formal education. For example, 30% firms pick higher focus scores ahead of any other skill or education but 10% rank resumes with higher focus skills last or second-last.

We interpret these smaller experimental findings as evidence that firms value the skills we measure and value better information about candidates' skills. This is consistent with the certification experiment and audit study implication that firms use information about skills in hiring.

4.5 'Skill-Blind' Certificates Have Limited Effects on Labor Market Outcomes

We interpret the certification as providing information about candidates' skills to the demand and supply sides of the labor market. However, they may work through a mechanism entirely unrelated to skill information. For example, certification may simply demonstrate that candidates were assessed, which firms may interpret as a positive signal about their skills or their dedication. Certification may instead change firms' job offer decisions by making applications more salient or visible. These two interpretations have different welfare implications, as better firm-worker matches are plausible under the former mechanism but not the latter mechanism. We implement a reduced-form test of the skill information interpretation against other interpretations but cannot distinguish between different special cases of the latter interpretation, like the dedication and salience ideas above.

We test these alternative interpretations by giving some candidates 'skill-blind' certificates that contain no information on assessment results but are otherwise identical to the public certificates (see Figure A.1). We randomly assign 254 candidates assessed over 3 days to this treatment arm. This is a deliberately small treatment arm because we judged that the skill information interpretation was more plausible and it was more importantly to have large public and private certification groups. The small sample size means that the treatment effect estimates are quite imprecise.

The skill-blind certification treatment has limited effects on labor market outcomes (Table A.9). The treatment effects are generally positive but on average only 25% as large as the public certification effects and not significantly different to zero. The small sample sizes mean that we generally cannot reject equality of the public and skill-blind certification effects. But the small magnitude of these treatment effects favors the skill information interpretation of our main results.

5 Conclusion

Firms make hiring decisions and workseekers make job search decisions based on potentially noisy signals of workseekers' skills and productivity. We argue that noisy signals can distort search and offer decisions, leading to lower employment and lower total earnings. This argument is particularly salient for populations with noisy skill signals such as young people with limited work experience and without educational qualifications that convey accurate information about their skills. With distortions at the employment margin, young workseekers may not get initial jobs that allow them to reveal their skills through experience.

We use a series of field experiments to validate this argument. Certifying skills to both the demand and supply sides of the labor market has large effects on employment and multiple employment quality measures. Treated candidates have 17% higher employment, 34% higher earnings, and 20% higher wages. These results show that certification gets more candidates into work and gets candidates into higher-paying jobs.

Our additional experiments directly show that frictions occur on both the supply and demand sides of the market. Revealing information to the supply side changes workseekers' beliefs and has modest effects on their job search and labor market outcomes. Revealing information to the demand side has imprecise but positive effects on the response rate to applications and interview invitation rate. This distinction is important. These findings suggest that actors on both sides of the market might be willing to pay for market-based provision of better information. Firms are indeed willing to pay for access to better information in a small incentivized choice experiment. Workseekers in our sample are willing to spend 17% of a hypothetical ZAR1,000 search subsidy on certification (compared to 24% on training and 27% on transport).

These findings raise several important questions for future work. We show that actors are willing to pay for access to better information; future work could explicitly test if market-based provision is financially viable. We show that skill certification information can change firms' hiring decisions; future work could examine if incorporating this information improves firm-worker matching algorithms or algorithmic hiring rules. We show that information on one specific set of skills is valuable but there are many other skills that could be assessed, whose value might differ substantially across sectors and occupations. Finally, our framework predicts and our results suggest that firm-level hiring and labor productivity will be lower in the presence of information frictions.

References

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge (2017): When Should You Adjust Standard Errors for Clustering? Working Paper 24003, National Bureau of Economic Research.

Abadie, A., M. Chingos, and M. West (2018): Endogenous Stratification in Randomized Experiments, *Review of Economics and Statistics*, 100, 567–580.

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2016): The Curse of Anonymity or Tyranny of Distance? The Impacts of Job-Search Support in Urban Ethiopia, Working Paper 22409, National Bureau of Economic Research.

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, S. Quinn, and F. Shilpi (2018): Job Fairs: Matching Firms and Workers in a Field Experiment in Ethiopia, Working paper, University of Oxford.

Abel, M., R. Burger, and P. Piraino (2019): The Value of Reference Letters: Experimental Evidence from South Africa, *American Economic Journal: Applied Economics*, forthcoming.

Acemoglu, D. and J.-S. Pischke (1999): Beyond Becker: Training in Imperfect Labour Markets, *The Economic Journal*, 109, F112–F142.

Alfonsi, L., O. Bandiera, V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2017): Tackling Youth Unemployment: Evidence from a Labor Market Experiment in Uganda, Manuscript, University College London.

Altmann, S., A. Falk, S. Jäger, and F. Zimmermann (2018): Learning about Job Search: A Field Experiment with Job Seekers in Germany, *Journal of Public Economics*, 164, 33–49.

Altonji, J., L. Kahn, and J. Speer (2015): Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success, Journal of Labor Economics, 34, 361 401.

Altonji, J. and C. Pierret (2001): Employer Learning and Statistical Discrimination, Quarterly Journal of Economics, 116, 313 335.

Anderson, M. (2008): Multiple Inference and Gender Di erences in the E ects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects, Journal of the American Statistical Association, 103, 1481 1495.

Arcidiaono, P., P. Bayer, and A. Hizmo (2010): Beyond Signaling and Human Capital: Education and the Revelation of Ability, American Economic Journal: Applied Economics, 2, 76 104.

Autor, D. and D. Scarborough (2008): Does Job Testing Harm Minority Workers? Evidence from Retail Establishments, Quarterly Journal of Economics, 123, 219 277.

Bartram, D. (2004): Assessment in Organisations, Applied Psychology: An International Review, 53, 237 259.

Bassi, V. and A. Nansamba (2017): Screening and Signaling Non-Cognitive Skills: Experimental Evidence from Uganda, Manuscript, University of Southern California.

Beam, E. (2016): Do Job Fairs Matter? Experimental Evidence from the Philippines, Journal of Development Economics, 120, 32 40.

Beaman, L., N. Keleher, and J. Magruder (2018): Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi, Journal of Labour Economics, 36, 121 153.

Beaman, L. and J. Magruder (2012): Who Gets the Job Referral? Evidence from a Social Networks Experiment. American Economic Review, 102, 3574 3593.

Belot, M., P. Kircher, and P. Muller (2018): Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice, The Review of Economic Studies, 86, 1411 1447.

Benjamini, Y., A. Krieger, and D. Yekutieli (2006): Adaptive Linear Step-Up Procedures That Control the False Discovery Rate, Biometrika, 93, 491 507.

Bertrand, M. and B. Crépon (2019): Teaching Labor Laws: Evidence from a Randomized Trial in South Africa, Manuscript, University of Chicago and CREST.

Bertrand, M. and S. Mullainathan (2004): Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, American Economic Review, 94, 991 1013.

Bhorat, H., T. Caetano, B. Jourdan, R. Kanbur, C. Rooney, B. Stanwix, I. Woolard, L. Patel, Z. Khan, L. Graham, K. Baldry, and T. Mqehe (2016): Investigating the Feasibility of a National Minimum Wage for South Africa, Manuscript, University of Johannesburg: Centre for Social Development in Africa and University of Cape Town: Development Policy Research Unit.

Bhorat, H. and H. Cheadle (20109): Labour Reform in South Africa: Measuring Regulation and a Synthesis of Policy Suggestions, Development Policy Research Unit Working Paper 09/139, University of Cape Town.

Blattman, C. and J. Annan (2016): Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State, American Political Science Review, 110, 1 17.

Bloom, N., A. Mahajan, D. McKenzie, and J. Roberts (2010): Why Do Firms in Developing Countries Have Low Productivity? The American Economic Review, 100, 619 623.

Botero, J., S. Djankov, R. L. Porta, F. Lopez-de Silanes, and A. Shleifer (2004): The Regulation of Labor, Quarterly Journal of Economics, 119, 1339 1382.

Budlender, J., M. Leibbrandt, and I. Woolard (2015): South African Poverty Lines: A Review and Two New Money-Metric Thresholds, Manuscript, Southern Africa Labour and Development Research Unit, University of Cape Town.

Burks, S., J. Carpenter, L. Goette, and A. Rustichini (2009): Cognitive Skills A ect Economic Preferences, Strategic Behavior, and Job Attachment, Proceedings of the National Academy of Sciences, 106, 7745 7750.

Card, D., J. Kluve, and A. Weber (2015): What Works? A Meta-Analysis of Recent Active Labor Market Program Evaluations, Working Paper 21431, National Bureau of Economic Research.

Clark, D. and P. Martorell (2014): The Signaling Value of a High School Diploma, Journal of Political Economy, 122, 282 318.

De Kock, F. and A. Schlechter (2009): Fluid Intellingence and Spatial Reasoning as Predictors of Pilot Training Performance in the South African Air Force (SAAF), SA Journal of Industrial Psychology, 35, 31 38.

Donovan, K., J. Lu, T. Schoellman, et al. (2018): Labor Market Flows and Development, in 2018 Meeting Papers, Society for Economic Dynamics, vol. 976.

Duckworth, A., C. Peterson, M. Matthews, and D. Kelly (2007): Grit: Perseverance and Passion for Long-term Goals, Journal of Personality and Social Psychology, 92, 1087 1101.

Eskreis-Winkler, L., A. L. Duckworth, E. Shulman, and S. Beal (2014): The Grit E ect: Predicting Retention in the Military, the Workplace, School and Marriage, Frontiers in Psychology, 5, 1 12.

Esopo, K., D. Mellow, C. Thomas, H. Uckat, J. Abraham, P. Jain, C. Jang, N. Otis, M. Riis-Vestergaard, A. Starcev, K. Orkin, and J. Haushofer (2018): Measuring Self-E cacy, Executive Function, and Temporal Discounting in Kenya, Behaviour Research and Therapy, 101, 30 45.

Fafchamps, M. and A. Moradi (2015): Referral and Job Performance: Evidence from the Ghana Colonial Army, Economic Development and Cultural Change, 63, 715 751.

Farber, H. and R. Gibbons (1996): Learning and Wage Dynamics, Quarterly Journal of Economics, 111, 1007 1047.

Foxcroft, C. (1997): Psychological Testing in South Africa: Perspectives Regarding Ethical and Fair Practices, European Journal of Psychological Assessment, 13, 229 235.

(2004): Planning a Psychological Test in the Multicultural South African Context, SA Journal of Industrial Psychology, 30, 8 15.

Freeman, R. (1999): The Economics of Crime, in Handbook of Labor Economics Volume 5, ed. by O. Ashenfelter and D. Card, Elsevier, 3529 3572.

García-Pérez, J. I., I. Marinescu, and J. Vall Castello (2018): Can Fixed-term Contracts Put Low Skilled Youth on a Better Career Path? Evidence from Spain, Economic Journal, 129, 1693 1730.

Garlick, R., K. Orkin, and S. Quinn (2019): Call Me Maybe: Experimental Evidence on Frequency and Medium E ects in Microenterprise Surveys, World Bank Economic Review, forthcoming.

Gneezy, U., A. Rustichini, and A. Vostroknutov (2010): Experience and Insight in The Race Game, Journal of Economic Behavior and Organization, 75, 144 155.

Groh, M., D. McKenzie, N. Shammout, and T. Viswanath (2015): Testing the Importance of Search Frictions and Matching through a Randomized Experiment in Jordan, IZA Journal of Labor Economics, 4.

Hall, R. and C. Jones (1999): Why Do Some Countries Produce So Much More Output per Worker than Others? Quarterly Journal of Economics, 114, 83 116.

Hardy, M. and J. McCasland (2017): Are Small Firms Labor Constrained? Experimental Evidence from Ghana, Working Paper, New York University - Abu Dhabi.

Heath, R. (2018): Why Do Firms Hire Using Referrals? Evidence from Bangladeshi Garment Factories, Journal of Political Economy, 126, 1691 1746.

Hepner, P. and C. Petersen (1982): The Development and Implications of a Personal Problem-Solving Inventory, Journal of Counseling Psychology, 29, 66 75.

Hoffman, M., L. Kahn, and D. Li (2018): Discretion in Hiring, Quarterly Journal of Economics, 133, 1 36.

Horton, J. (2017): The E ects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment, Journal of Labor Economics, 35, 345 285.

Hsieh, C.-T. and P. J. Klenow (2014): The Life Cycle of Plants in India and Mexico *, The Quarterly Journal of Economics, 129, 1035 1084.

Hsieh, C.-T. and B. Olken (2014): The Missing `Missing Middle', Journal of Economic Perspectives, 28, 89 108.

ILO (2016): Enabling Environment for Sustainable Enterprises in South Africa, International Labour O ce Geneva.

Ioannides, Y. M. and L. D. Loury (2004): Job Information Networks, Neighborhood E ects, and Inequality, Journal of Economic Literature, 42, 1056 1093.

Isaacs, G. (2016): A National Minimum Wage for South Africa, Manuscript, University of the Witwatersrand: National Minimum Wage Research Initiative.

Kahn, L. (2010): The Long-term Labor Market Consequences of Graduating from College in a Bad Economy, Labour Economics, 17, 303 316.

(2013): Asymmetric Information between Employers, American Economic Journal: Applied Economics, 5, 165 205.

Kahn, L. and F. Lange (2014): Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures, Review of Economic Studies, 84, 1575 1613.

Kessler, J., C. Low, and C. Sullivan (2019): Incentivized Resume Rating: Eliciting Employer Preferences without Deception, American Economic Review, forthcoming.

Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2018): Life Cycle Wage Growth Across Countries, Journal of Political Economy, 126, 797 849.

Lam, D., C. Ardington, and M. Leibbrandt (2011): Schooling as a Lottery: Racial Di erences in School Advancement in urban South Africa, Journal of Development Economics, 95, 121 136.

Lange, F. (2007): The Speed of Employer Learning, Journal of Labor Economics, 25, 1 35.

Leibbrandt, M., A. Finn, and I. Woolard (2012): Describing and Decomposing Post-Apartheid Income Inequality in South Africa, Development Southern Africa, 29, 19 34.

Levinsohn, J., N. Rankin, G. Roberts, and V. Schoer (2013): Wage Subsidies and Youth Employment in South Africa: Evidence from a Randomized Control Trial, Manuscript, University of Stellenbosch.

Lopes, A., G. Roodt, and R. Mauer (2001): The Predictive Validity of the APIL-B in a Financial Institution, SA Journal of Industrial Psychology, 27, 61 69.

Mackay, A. (2014): Building Brands in a Rapidly Changing Market: Lessons for South Africa, Yellowwood.

MacLeod, W. B., E. Riehl, J. Saavedra, and M. Urquiola (2017): The Big Sort: College Reputation and Labor Market Outcomes, American Economic Journal: Applied Economics, 9, 223 261.

Magruder, J. (2010): Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa, American Economic Journal: Applied Economics, 2, 62 85.

Neumark, D. (2018): Experimental Research on Labor Market Discrimination, Journal of Economic Literature, 56, 799 866.

Oreopoulos, P., T. Von Wachter, and A. Heisz (2012): The Short-and Long-Term Career E ects of Graduating in a Recession, American Economic Journal: Applied Economics, 4, 1 29.

Pallais, A. (2014): Ine cient Hiring in Entry-Level Labor Markets, American Economic Review, 104, 3565 3599.

Paterson, H. and K. Uys (2005): Critical Issues in Psychological Test Use in the South African Workplace, South African Journal of Industrial Psychology, 31, 12 22.

Pritchett, L. (2013): The Rebirth of Education: Schooling Ain't Learning, Washington, DC: Center for Global Development.

Rankin, N., C. Darroll, and T. Corrigan (2012): SMEs and Employment in South Africa, Small Business Project, Johannesburg.

Raven, J. and Raven, J. (2003): Raven Progressive Matrices, in Handbook of Nonverbal Assessment, ed. by R. McCallum, Boston: Springer, 223 237.

Riach, P. and J. Rich (2002): Field Experiments of Discrimination in the Market Place, Economic Journal, 112, F480 518.

Rich, J. (2014): What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000, Discussion Paper 8584, Institute for the Study of Labor (IZA).

Rosenberg, M. (1965): Society and the Adolescent Self-Image, Princeton University Press.

Schmidt, K., B. Neubach, and H. Heuer (2007): Self-Control Demands, Cognitive Control De cits, and Burnout, Work and Stress, 21, 142 154.

Schöer, V., M. Ntuli, N. Rankin, C. Sebastiao, and K. Hunt (2010): A Blurred Signal? The Usefulness of National Senior Certi cate (NSC) Mathematics Marks as Predictors of Academic Performance at University Level, Perspectives in Education, 28, 9 18.

Schöer, V., N. Rankin, and G. Roberts (2014): Accessing the First Job in a Slack Labour Market: Job Matching in South Africa, Journal of International Development, 26, 1 22.

Small Business Project (2013): Easier, Harder for Small Business in South Africa: Headline Report of SBP's SME Growth Index, Small Business Project, Johannesburg.

Söderbom, M. and F. Teal (2004): Size and E ciency in African Manufacturing Firms: Evidence from Firm-level Panel Data, Journal of Development Economics, 73, 369 394.

Statistics South Africa (2013): Gender Statistics in South Africa, Statistics South Africa, Pretoria.

(2016): Quarterly Labor Force Survey Quarter 4 2016, Statistics South Africa, Pretoria.

Stroop, J. R. (1935): Studies of Interference in Serial Verbal Reactions, Journal of Experimental Psychology, 18, 643 662.

Taylor, S., S. Van Der Berg, V. Reddy, and D. Janse van Rensburg (2011): How Well Do South African Schools Convert Grade 8 Achievement Into Matric Outcomes? Tech. Rep. 13/11, Stellenbosch Economic Working Papers.

Taylor, T. (1994): A Review of Three Approaches to Cognitive Assessment, and a Proposed Integrated Approach Based on a Unifying Theoretical Framework, South African Journal of Psychology, 24.

(2013): APIL and TRAM Learning Potential Assessment Instruments, in Psychological Assessment in South Africa, ed. by S. Laher and K. Cockroft, Wits University Press, 158 168.

Tourangeau, R. (2003): Cognitive Aspects of Survey Measurement and Mismeasurement, International Journal of Public Opinion Research, 15, 3 7.

Van der Berg, S. and D. Shepherd (2015): Signalling Performance: Continuous Assessment and Matriculation Examination Marks in South African Schools, South African Journal of Childhood Education, 5, 78 94.

Wheeler, L., E. Johnson, R. Garlick, P. Shaw, and M. Gargano (2019): LinkedIn(to) Job Opportunities? Experimental Evidence from Job Readiness Training, Working paper, Duke University.

Willis, G. B. (1999): Reducing Survey Error Through Research on the Cognitive and Decision Processes in Surveys, Short Course Presented at the 1999 Meeting of the American Statistical Association.

(2008): Cognitive Aspects of Survey Methodology (CASM), in Sage Encyclopedia of Survey Research Methods, ed. by P. J. Lavrakas, Thousand Oaks, California: Sage Publications, 104 106.

Witte, M. (2019): Job Referrals and Strategic Network Formation: Experimental Evidence from Urban Neighbourhoods in Ethiopia, Manuscript, University of Oxford.

World Bank (2018): World Development Report 2018: Learning to Realize Education's Promise, World Bank.

World Economic Forum (2018): The Future of Jobs Report, World Economic Forum Geneva.

A Further Details on Skills Assessments

A.1 Skills Measures

We assess workseekers' skills in six domains. Detailed information on all six assessments is available at https://www.assessmentreport.info , including sample questions. The numeracy, literacy and concept formation assessments are registered with the South African Qualifications Authority and widely in use. All assessments are conducted on desktop computers, so the assessment results will be driven in part by candidates' computer skills.

Concept formation is a non-verbal measure of fluid intelligence and captures conceptual reasoning the ability to ignore superficial differences and see underlying commonalities and to use logic in new situations. The CFT adopts a multiple choice, non-verbal approach and suitable for candidates who do not speak English as a first language. It avoids using any curriculum-based knowledge. It is a subtest from the TRAM 2, which is normed for South Africans with 10 to 12 years of learning (Taylor, 1994). It is similar in approach to the Ravens Progressive Matrices (Raven, J. and Raven, J., 2003). It was correlated with interview ratings and technical scores in Johannesburg Municipality clerks and supervisor ratings for administrative clerks at a phone company and an import-export firm (Taylor, 2013). The full battery predicted performance in a financial institution and pilot training (Lopes et al., 2001; De Kock and Schlechter, 2009).

Numeracy focuses on practical arithmetic and pattern recognition. We calculate a single numeracy score using the inverse variance-weighted average of two numeracy assessment scores. The more advanced assessment is developed by a large retail chain and used in their applicant screening process. The assessments evaluate candidates' ability to compare different types of numbers, to work with fractions, ratios, money, percentages and units, and to perform calculations with time and area. Literacy and communication captures English language listening, reading and comprehension skills by testing comprehension of spoken and written passages. These two assessments were developed for candidates of a similar age and education range in the South African context by a South African adult education provider (https://www.mediaworks.co.za/). They capture where a candidate places between Level 1 and Level 4 on the National Qualifications Framework, corresponding roughly to Grade 9 to matric.

We then developed adapted three other assessments for the context from measures used by psychologists and behavioral economists in other contexts:

- Grit is a self-reported measure of a candidate's inclination to work on difficult tasks until they are finished and whether they show perseverance to achieve long-term goals. We use the 8 item measure from Duckworth et al. (2007). Grit correlates with academic performance and workplace retention (Eskreis-Winkler et al., 2014).

- Focus measures a candidate's ability to distinguish relevant from irrelevant information in potentially confusing environments. Our assessment is a shortened and computerized version of the widely-used Stroop Test, using colors (Stroop, 1935). Similar characteristics to those measured by the Focus Test have been shown to moderate the negative effects of workplace related stress such as burnout and absenteeism in service sector jobs in Germany (Schmidt et al., 2007).

Planning measures how candidates behave when faced with complex, multi-step problems. Our assessment is adapted from a test proposed by Gneezy et al. (2010) called the Hit 15 task. The computer and the subject take turns adding points to the points basket and in each turn the subject or the computer must add either one, two, or three points to the points basket. The goal is to be the first player to reach 15 points. High planning ability has been shown to predict retention rates among truckers in the US, even when other factors, such as IQ, are taken into account (Burks et al., 2009).

For the first 23% of the workseekers we assessed, we used self-reported measures of **control** and **flexibility** instead of the focus and planning assessments. We use two subscales of the Personal Problem-Solving Inventory Hepner and Petersen (1982). The control scale is a self-reported measure of whether candidates take a systematic or impulsive and erratic approach when faced with new, challenging problems. The flexibility scale is a self-reported scale which captures whether candidates actively consider several approaches to solving a problem or whether they pursue their first idea without thinking about alternatives.

To validate the scales, we conducted cognitive debriefings with 20 Harambee participants. Cognitive debriefing captures the underlying cognitive processes that respondents use to answer questions to detect and solve problems in questionnaires (Tourangeau, 2003; Willis, 2008, 1999). For example, the interviewer asks for specific information relevant to the question or the answer given. Examples of probes used are "What does the term mean to you?", "Can you repeat this question to me in your own words?" and "What made you answer the way that you did?" After these cognitive debriefings, changes to the wording of some items were made.

With 150 respondents, we administered the tests twice ten days apart. We used this dataset to conduct standard psychometric validation on the scales (Esopo et al., 2018).

A.2 Firms' Use of Assessments

A.3 Further Details on Process

Verification of Eligibility: In the phone screening, candidates are told they may be asked to provide documented proof of their answers to screening questions. Before the first day of assessments start, candidates are reminded of the criteria and asked to leave the room to meet with a psychologist if they are not eligible. The psychologist verifies their eligibility and they leave the assessments if they are not eligible.

Table A.1: Firms' Use of Psychometric Assessments in Hiring

		# Harambee client firms using each score or piece of information to screen candidates								
Industry	# firms	Literacy & comms.	CFT	Basic numeracy	Advanced numeracy	Soft skills	Crim. check	Matric results	Reference	CV with Reference
Hospitality	11	9	11	10	7	7	10	7	0	9
%		0.82	1.00	0.91	0.64	0.64	0.91	0.64	0.00	0.82
Retail	16	11	9	7	14	13	15	12	1	5
%		0.69	0.56	0.44	0.88	0.81	0.94	0.75	0.06	0.31
Corp.	6	6	6	6	5	2	6	0	0	6
%		1.00	1.00	1.00	0.83	0.33	1.00	0.00	0.00	1.00
Total	33	26	26	23	26	22	31	19	1	20
%		0.79	0.79	0.70	0.79	0.67	0.94	0.58	0.03	0.61

This table captures use of psychometric and other assessment scores by 33 Harambee client firms. The assessments are described in Appendix A. Firms coded as using an assessment required candidates to reach a certain threshold score on the assessment to be eligible for interviews or training programs. Firms coded as requiring other documents required these to be submitted with the candidate's application package but we do not know how these were used. The `criminal' check was a set of checks against government records that the candidate had no criminal record or bad credit history and had actuallly passed matric.

On the second day of assessments, candidates have to provide certified copies of their identity document and matric certificates. Harambee does not formally verify economic status, employment history or status, or school enrolment. Criminal record and credit checks are only completed later in the process for the small portion of candidates who are selected to be matched to employers.

Compliance With Legal Requirements on Psychometric Testing: In South Africa, only registered psychologists are permitted to administer and report on psychometric or personality tests (see https://www.hpcsa.co.za/Uploads/editor/UserFiles/downloads/psych/psycho_policy/form_208.pdf). A Harambee psychologist approved the design of certificates. Psychologists oversaw each testing session and delivered briefings to candidates to interpret results.

Sensitivity to Candidate Circumstances: Tests are split over two days to avoid tiredness. Candidates receive breakfast and lunch on both assessment days. Harambee has a specially designed assessment stream tailored for people living with disabilities. Psychologists and a specialized team of facilitators are trained to work with candidates with disabilities. Candidates who report a disability have an assessment with a psychologist on the first day of assessments to understand the nature of the disability. Candidates with disabilities related to mobility, some psychological disabilities, and some learning disabilities can be accommodated by Harambee and sit assessments using a medium that accommodates their needs. At the time of our experiment, Harambee did not have provisions for candidates with hearing or visual impairments. Candidates with disabilities are not included in our sample.

B Decomposing Labor Market Effects into Extensive and Intensive Margins

Treatment effects on labor market outcomes such as earnings and hours can occur at the extensive margin due to treatment effects on employment and at the intensive margin due to treatment effects on job characteristics conditional on employment. This distinction is important, as intensive margin effects indicate that treatment is changing the type of jobs candidates secure. The intensive margin effects are not identified from regressions of labor market outcomes on treatment indicators for employed candidates, as set of employed candidates may be selected based on treatment assignment.

We propose a simple but, to the best of our knowledge, novel decomposition of labor market effects into extensive and intensive margins. We describe the decomposition here for earnings but the same idea applies to any labor market outcome that is observed only for the employed. Using the law of iterated expectations and the fact that observed earnings are zero for non-employed candidates, we can write the average treatment effect on earnings as:

$$\begin{aligned}
&\underbrace{E[Earn|Treat=1]-E[Earn|Treat=0]}_{\text{ATE for earnings}} \qquad (6)\\
&= \left(\underbrace{E[Earn|Treat=1,Work=1]-E[Earn|Treat=0,Work=1]}_{\text{ATE for earnings}\,|\,\text{employment}}\right)\times\underbrace{Pr[Work=1|Treat=1]}_{\text{Treated employment rate}}\\
&+\underbrace{E[Earn|Treat=0,Work=1]}_{\text{Control earnings}\,|\,\text{employment}}\times\left(\underbrace{Pr[Work=1|Treat=1]-Pr[Work=1|Treat=0]}_{\text{ATE for employment}}\right).
\end{aligned}$$

We define the second line on the right-hand of the regression as the extensive margin effect. Intuitively, this is the average treatment effect on employment 'priced' at the mean earnings value in the control group. If treatment has no effect on the employment rate, then this expression is zero. We define the first line on the right-hand side of the regression as the intensive margin effect. If treatment only changes the employment rate but has no effect on earnings for employed candidates, then this term is zero.

All terms in equation (6) except the average treatment effect on earnings conditional on employment are identified by the experiment and can be consistently estimated using sample analogues. Hence we can consistently estimate the remaining term using the formula in (6). We obtain standard errors by estimating all quantities as a system and using the Delta method.

This decomposition applies to *observed* earnings, which are zero by definition for non-employed candidates. This decomposition does not apply to *latent* earnings, which may be non-zero for non-employed candidates. We could alternatively study average treatment effects on latent earnings using a selection correction model or quantile treatment effects on latent earnings by assuming that latent earnings for non-employed workers are below some percentile of the observed earnings distribution. The latter approach has been used in evaluations of active labor market programs in some settings. But it is not attractive in settings like this with low employment rates, where earnings are unobserved for most candidates.

As discussed in Section 3.3, this decomposition shows that the earnings effects of skill certification occur at both the extensive and intensive margins. The hours and contract type effects occur only at the extensive margin.

C Heterogeneous Outcome Levels and Treatment Effects by Skill

Analyzing the relationship between `skill' and other candidate attributes, including treatment effects, is difficult because we observe multiple measures of skill. In this appendix we explain the challenge of estimating heterogeneous treatment effects by skill and a new test we propose to address this. We conclude by showing that a special case of this test can be used to establish if baseline levels of employment, education, etc. differ by skill.

There are two obvious parametric approaches to testing for heterogeneous treatment effects by skill, each of which has serious limitations. First, we could collapse multidimensional skills into a single index and then regress outcomes on treatment assignments, the single skill index, and their interaction. However, this imposes the assumption that any treatment effect heterogeneity over skills is linear and additive, ruling out the possibility of complementarity between different skills. Second, we could regress outcomes on treatment assignments, each skill measure, and all possible interactions. However, this requires estimating a large number of parameters and may require imposing some rule to aggregate over potentially different signs.

We instead use a nonparametric test based on the idea of skill dominance. To build intuition for this test, consider the set of candidates scoring in the middle tercile for all skill assessments. Let the average treatment effect for this subset of candidates be τ_M. If treatment effects are increasing in skill, then τ_M should be larger than the average treatment effect for candidates scoring in the lowest tercile for all skill tests, τ_L. Similarly, τ_M should be smaller than the average treatment effect for candidates scoring in the highest tercile for all skill tests, τ_H. More generally, the average treatment effect for candidates with any skill level should be higher (respectively lower) than the average treatment effect for candidates with strictly lower (respectively higher) skills in all domains.

More formally, define

$$\tau_{s_1,\dots,s_J} = E\left[Y_i(1) - Y_i(0) \mid S_{i,1} = s_1, \dots, S_{i,J} = s_J\right] \tag{7}$$

as the average treatment effect for candidates scoring in terciles $s_1, \dots, s_J$ for the J assessments. This expectation is taken over the distibution of candidate-level heterogeneity conditional on skill. Define

$$\tau^{+}_{s_1,\dots,s_J} = \tau_{s_1,\dots,s_J} - E\left[\tau_{t_1,\dots,t_J}\right] \tag{8}$$

equal the difference in the average treatment effect between candidates with skills $s_1, \dots, s_J$ and candidates with skills $t_1, \dots, t_J$, where $t_j < s_j \; \forall j$. The expectation in the second term of the right-hand side expression is taken over the distribution of all skills that are strictly lower than s_j in all dimensions j. Define

$$\tau^{-}_{s_1,\dots,s_J} = \tau_{s_1,\dots,s_J} - E\left[\tau_{t_1,\dots,t_J}\right] \tag{9}$$

analogously, where $t_j > s_j \; \forall j$.

If treatment effects are increasing in skill, then $\Delta^{+}_{s_1,\ldots,s_J} > 0 > \Delta^{-}_{s_1,\ldots,s_J}$. The signs of the two terms do not depend on the sign of the average treatment effect because both are deviations from the sample average treatment effect. This motivates the three hypothesis tests

$$H_{+}: \Delta^{+}_{s_1,\ldots,s_J} = 0$$
$$H_{-}: \Delta^{-}_{s_1,\ldots,s_J} = 0$$
$$H_{d}: \Delta^{+}_{s_1,\ldots,s_J} - \Delta^{-}_{s_1,\ldots,s_J} = 0.$$

We use two-sided tests because treatment effects may be increasing or decreasing in skill.

We implement this test by estimating the average treatment effect within each skill cell and the differences between average treatment effects using respectively sample means and differences between sample means. We estimate standard errors using two-way clustering by baseline date (the unit of treatment assignment) and skill cell (as the averages used in the test only vary at this level). Simulations calibrated to our data show that tests based on two-way clustered standard errors have approximately the correct size, but can substantially overreject the null of no treatment effect heterogeneity when the number of skills is lower than in our measures.

As discussed in Section 3.3 and reported in Table 3, we do not find strong evidence of heterogeneous skill certification effects on labor market outcomes by skill. Certification effects on employment and hours are slightly higher for candidates with lower skills. Certification effects on earnings, wages, and contract status do not vary by skill. The strongest evidence of certification effect heterogeneity is for search with certificates. Extensive margin certificate use, applications with certificates, interviews from applications with certificates, and job offers from applications with certificates are all increasing in skill.

This test imposes no assumptions about the relative magnitude of treatment effects across different skills or about complementarity between different skills. But it does assume that treatment effects across different skills have the same sign. Consider a concrete but extreme example. First, assume certification shifts firms from having zero information about candidates skills to full information, that all firms value numeracy but dislike communication skills, that there are no other search or matching frictions, and that numeracy and communication skills are independent across candidates. Then the dominance test will average over positive heterogeneity by numeracy and negative heterogeneity by communication and conclude that treatment effects do not vary by skill. To address this possibility, we also implement two parametric heterogeneity tests.

There is little evidence of heterogeneous skill certification effects on labor market outcomes using tests based on regressions with treatment-skill interactions. We estimate models of the form

$$Y_{id} = T_d\beta + \sum_{j=1}^{J} \text{Skill}^{j}_{id} \cdot T_d \cdot \gamma_j + X_{id}\delta + S_d + \varepsilon_{id}, \tag{10}$$

where Skill^{j}_{id} is either the standardized score or tercile for candidate i in skill j. The estimates of $_j$ are generally small and statistically insigni cant. In particular, we see no strong evidence of heterogeneous e ects with di erent signs over di erent skills, which the dominance test would fail to detect.

There is also little evidence of heterogeneous skill certi cation e ects by the variance of skills on labor market outcomes. We estimate models of the form

$$Y_{id} = T_d \quad + Skill^M_{id} \; T_d \quad _M + Skill^D_{id} \; T_d \quad _D + X_{id} \quad + \; S_d + \; _{id}; \tag{11}$$

where $Skill^M_{id}$ is a summary measure for mean skill across all dimensions (rst principal component of standardized scores, inverse covariance-weighted average of standardized scores, or index constructed by assigning 1 for each middle tercile and 2 for each top tercile) and $Skill^D_{id}$ is a summary measure for the dispersion of skill across all dimensions (range across terciles, number of top terciles, variance across terciles, variance across standardized scores). The estimates of both $_M$ and $_D$ are generally small and statistically insigni cant. This provides evidence against a model where rms value specialized candidates with high values of just one skill.

We can use this dominance approach to test if the level of baseline variables di er by skill. We simply de ne

$$Y_{s_1;:::;s_J} = E[Y_i | S_{i;1} = s_1; : : : ; S_{i;J} = s_J] \tag{12}$$

instead of $_{s_1;:::;s_J}$ and then de ne $Y_{s_1;:::;s_J}$ and $Y^+_{s_1;:::;s_J}$ analogously.

D Additional Results Discussed in Paper

D.1 Summary Statistics and Balance Tests

This section reports summary statistics for the baseline workseeker sample (Table A.2), endline workseeker sample (Table A.3), and sample of workseekers used in the audit study (Table A.4). Balance tests for equal means of baseline measures are also reported in the nal column of Table A.2.

Table A.2: Summary Statistics for Baseline Variables

Variable	# obs	Mean	Std dev.	10^{th} pctile	90^{th} pctile	p:balance
Panel A: Demographic Measures						
Age	6891	23.6	3.3	19.8	28.3	0.584
Male	6891	0.381	0.486			0.275
University degree	6889	0.167	0.373			0.887
Any other post-secondary quali cation	6889	0.212	0.409			0.644
Completed secondary education only	6891	0.610	0.488			0.790
Panel B: Assessment Results						
Numeracy score	6891	0.052	0.988	-1.187	1.411	0.514
Communication score	6891	0.050	0.992	-1.093	1.694	0.206
Concept formation score	6891	0.047	0.991	-1.516	1.260	0.769
Grit score	6891	0.031	0.992	-1.313	1.279	0.088
Other scores	6891	-0.002	1.070	-1.298	1.318	0.862
Panel C: Labor Market Measures						
Worked in past 7 days	6891	0.378	0.485			0.459
Earnings in past 7 days	2116	560	712	100	1400	0.105
Panel D: Job Search Measures						
Any search in past 7 days	6891	0.968	0.176			0.058
Applications submitted in past 30 days	6813	9.4	13.5	2.0	20.0	0.812
Search cost in past 7 days	6145	189	215	30	400	0.721
Search hours in past 7 days	6697	16.7	19.6	2.0	48.0	0.201
Responses received in past 30 days	6772	0.743	1.377	0.000	2.000	0.396
O ers received in past 30 days	6809	0.822	2.722	0.000	2.000	0.840
Panel E: Belief Measures						
Planned applications in next 30 days	6838	17.8	25.5	4.0	36.0	0.127
Self-esteem index	6891	5.38	1.15	3.80	6.80	0.545
Fraction of assessments overcon dent	6873	0.503	0.352			0.588
Fraction of assessments undercon dent	6873	0.115	0.208			0.367

Table shows summary statistics for selected baseline variables. Percentiles are omitted for binary variables. All monetary gures are reported in South Africa Rands. 1 Rand USD0.16 in purchasing power parity terms. All assessment results are standardized to have mean zero and standard deviation one in the control group. Intensive-margin labor market measures are set to missing for non-workers. Intensive-margin search measures are winsorized at the 99th percentile and set to zero for non-searchers. Final column reports the p-value for testing equality of means of the baseline variables across all treatment groups.

Table A.3: Summary Statistics for Endline Variables

Variable	# obs	Mean	Std dev.	10^{th} pctile	90^{th} pctile
Panel A: Labor Market Measures					
Worked in past 7 days	6605	0.323	0.468		
Earnings in past 7 days	2112	623	1183	2	1500
Hours worked in past 7 days	2121	28.5	21.6	4.0	56.0
Hourly wage	2097	33.1	72.3	0.1	77.8
Written contract	2100	0.401	0.490		
Written permanent contract	2100	0.035	0.183		
Wage employment	2100	0.563	0.496		
Self employment	2100	0.251	0.434		
Family employment	2100	0.114	0.318		
Panel B: Job Search Measures					
Any search in past 7 days	6606	0.692	0.462		
Applications submitted in past 30 days	6860	12.3	21.2	0.0	25.0
Search hours in past 7 days	6599	11.2	14.0	0.5	27.0
Search cost in past 7 days	6597	129	163	0	300
Responses received in past 30 days	6593	0.861	2.147	0.000	2.000
O ers received in past 30 days	6875	0.198	0.667	0.000	1.000
Panel C: Belief Measures					
Planned applications in next 30 days	6589	14.9	19.0	3.0	30.0
Fraction of assessments overcon dent	6605	0.345	0.237		
Fraction of assessments undercon dent	6605	0.176	0.166		

Table shows summary statistics for selected endline variables. Percentiles are omitted for binary variables. All monetary gures are reported in South Africa Rands. 1 Rand USD0.16 in purchasing power parity terms. Intensive-margin search measures are winsorized at the 99th percentile and set to zero for non-searchers. Intensive-margin labor market measures are set to missing for non-workers.

Table A.4: Summary Statistics for Workseekers Participating in the Audit Study

	(1)	(2)	(3)	(4)	(5)	(6)
	Audit sample			Workseeker sample		
	Mean	Std dev.	# obs	Mean	Std dev.	# obs
Panel A: Characteristics of Applications Received from Workseekers						
Includes a cover letter	0.13	0.34	492	-	-	
Includes a copy of ID document	0.50	0.50	630	-	-	
Includes a drivers license	0.12	0.32	630	-	-	
Includes information about Matric	0.59	0.49	630	-	-	
Includes references or a reference letter	0.90	0.30	628	-	-	
Above median quality score	0.51	0.50	617	-	-	
Panel B: Characteristics of Workseekers						
Public group	0.31	0.46	632	0.33	0.47	6,891
Private group	0.37	0.48	632	0.31	0.46	6,891
Placebo group	0.00	0.00	632	0.04	0.19	6,891
Age	23.29	3.15	632	23.65	3.30	6,891
Male	0.48	0.50	632	0.38	0.49	6,891
Completed diploma or degree	0.18	0.39	632	0.17	0.37	6,891
Completed post-high-school certi cate	0.24	0.43	632	0.21	0.41	6,891
Completed high-school	0.57	0.50	632	0.61	0.49	6,891
Completed less than high-school	0.43	0.50	632	0.39	0.49	6,891
Numeracy assessment score (z-score)	0.05	0.96	632	0.05	0.99	6,891
Literacy assessment score (z-score)	-0.01	0.94	632	0.05	0.99	6,891
Concept Formation assessment score (z-score)	0.11	0.92	632	0.05	0.99	6,891
Worked in the last 7-days (endline)	0.41	0.49	632	0.38	0.48	6,891

D.2 Benchmarks for Earnings Figures

Minimum wage : A national minimum wage was only instituted in January 2019. Before this, minimum wages were either set by sector by the Ministry of Labour or in bargaining councils, where large rms and unions agreed minimum wages for the sector which were often then applied to all rms in the sector (Budlender et al., 2015; Isaacs, 2016). Table A.5 shows a few minimum wages for urban areas for low-skilled occupations at the time of the baseline.

Poverty Lines : South African poverty research often uses the upper poverty line (the food poverty line to get 2100 calories plus the average amount spent on non-food items by households whose food expenditure equals the food poverty line (Leibbrandt et al., 2012). Lines for 2011 were calculated using these methods from nationally representative survey data (Budlender et al., 2015). They were then updated to February 2016, six months before the start of the baseline, the poverty line was 1386 ZAR per adult per month (Isaacs, 2016, p.22). Household poverty lines are calculated assuming one working adult per household of four, the median composition of South African households. The household poverty line would be 5544 ZAR.

Table A.5: Benchmarking Earnings Figures to Minimum Wage and Poverty Lines

Panel A: South African poverty lines and minimum wages at baseline

		Monthly		Weekly	
	Date	ZAR	USD	ZAR	USD
Poverty line					
Adult upper	Early 2016	1386	222	308	49
Household upper (4 people)	Early 2016	5544	887	1232	197
Minimum wage					
Domestic work	2015-2016	2550	408	567	91
Hospitality	2015-2016	2750	440	611	98
Wholesale and retail	2015-2016	3250	520	722	116
Private security/contract cleaning	2015-2016	3500	560	778	124

Panel B: Benchmarking sample earnings and certi cation treatment e ects on earnings

		Weekly		As % of poverty line		As % of min. wage	
Endline	Date	ZAR	USD	Adult	Household	Hospitality	Retail
Mean earnings	Early 2017	159.36	25	0.52	0.18	0.26	0.22
Mean earnings if employed	Early 2017	518.29	83	1.68	0.58	0.85	0.72
Treatment e ect	Early 2017	53.86	9	0.17	0.06	0.09	0.07
Baseline							
Mean earnings if employed	Late 2016	559.9	90	1.82	0.63	0.92	0.78

Calculations assume 1 Rand 0.16 USD in purchasing power parity terms; 4.5 weeks per month. Household poverty lines assume households of four people with only one earner. Note that control group respondents work 29 hours per week conditional on being employed; earnings for those in full time work will be higher than mean earnings here. Poverty lines are from Isaacs (2016, p.22); minimum wages are from Isaacs (2016, p.22) from the Department of Labor for 2015. Minimum wages are for large urban areas (Area A), grade D security guards, hospitality businesses with less than 10 employees, and shop assistants in the wholesale and retail sector.

D.3 Non-response

The phone survey after 3-4 months is our main source of endline data. We use the text message survey after 2-3 days only to measure beliefs about numeracy and self-esteem. The response rates for the text message and phone surveys are respectively 83 and 96%. Non-response does not di er by treatment arm (Table A.6). Non-response does not di er over most baseline characteristics. Men are less likely to respond in both surveys. Higher numeracy and concept formation scores predict higher response rates in the text message survey. Higher grit predicts lower response rates in the endline survey.

Table A.6: Non-response by Treatment Group in Each Post-Treatment Survey Round

	(1) Text Message Survey	(2) Endline Phone Survey
Control	0.170	0.040
	(0.013)	(0.006)
Private	0.182	0.044
	(0.010)	(0.004)
Public	0.178	0.039
	(0.012)	(0.004)
Placebo	0.142	0.047
	(0.032)	(0.026)
p: Control = Pvt.	0.484	0.634
p: Control = Pub.	0.670	0.855
p: Pvt. = Pub.	0.789	0.389
p: Control = Pvt. = Pub.	0.781	0.683
p: Control = Plc.	0.413	0.788
p: Pvt. = Plc.	0.238	0.888
p: Pub. = Plc.	0.296	0.747
p: Control = Pvt. = Pub. = Plc.	0.642	0.843
# observations	6889	6889
# clusters	84	84

Coe cients show the fraction of each treatment group that does not complete each follow-up survey round. Heteroskdasticity-robust standard errors clustered by treatment date are shown in parentheses.

Table A.7: Non-response by Baseline Covariates Group in Each Post-Treatment Survey Round

	(1) Text Message Survey	(2) Endline Phone Survey
Numeracy score	-0.029	0.003
	(0.006)	(0.003)
Communication score	0.009	0.004
	(0.006)	(0.003)
Concept formation score	-0.019	0.002
	(0.006)	(0.003)
Grit score	-0.002	-0.007
	(0.005)	(0.003)
Other scores	0.001	-0.002
	(0.004)	(0.003)
Perceived numeracy score	-0.000	-0.000
	(0.000)	(0.000)
Perceived literacy score	0.014	-0.003
	(0.010)	(0.004)
Perceived concept formation score	0.010	-0.003
	(0.009)	(0.004)
Self-esteem index	0.006	0.002
	(0.004)	(0.002)
Completed at most high school	-0.008	-0.003
	(0.012)	(0.005)
Age	-0.002	0.001
	(0.001)	(0.001)
Male	0.048	0.014
	(0.010)	(0.005)
Worked in last 7 days	-0.005	-0.001
	(0.008)	(0.005)
p: All coe cients jointly zero	0.000	0.018
Mean outcome	0.170	0.040
# observations	6889	6889
# clusters	84	84

Coe cients are from regressions of round-speci c attrition on the list of baseline covariates displayed here. All assessment scores are standardized to have mean zero and standard deviation one in the control group. Heteroskdasticity-robust standard errors clustered by treatment date are shown in parentheses.

D.4 Additional Treatment E ects

Table A.8: Treatment E ects on Beliefs about Skills, Self-Esteem and Returns to Search

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	# terciles correct	Numeracy belief correct	Numeracy belief correct	Self-esteem	Self-esteem	Expected search time	Expected o ers
	Endline			Endline			
	Endline		Post-treat		Post-treat	Endline	
Public	0.951	0.233	0.316	0.001	-0.001	-0.038	0.107
	(0.048)	(0.013)	(0.015)	(0.013)	(0.015)	(0.013)	(0.019)
Private	0.741	0.200	0.333	-0.002	0.016	-0.024	0.054
	(0.047)	(0.015)	(0.016)	(0.015)	(0.015)	(0.015)	(0.023)
p: pub. = priv	0.000	0.010	0.248	0.811	0.236	0.288	0.024
Mean outcome	2.335	0.396	0.399	0.553	0.479	0.408	1.807
# observations	6605	6599	5295	6607	5026	6330	6529
# clusters	84	84	84	84	84	84	84

Coe cients are from regressing each outcome on a vector of treatment assignments, randomization block xed e ects, and prespeci ed baseline covariates (measured skills, self-reported skills, education, age, gender, employment, discount rate, risk aversion). Heteroskedasticity-robust standard errors shown in parentheses, clustering by treatment date. Mean outcome is for the control group. `Post-treat' columns denote variables measured 2-3 days post-treatment. Endline columns denote the same variables measured at endline. `# terciles correct' is a score out of 6 measuring on how many skills candidates correctly report the tercile they are in. The numeracy belief measure captures whether candidates correctly report which tercile they are in for the numeracy test. This is the only skill we measure beliefs about both post-treatment and at endline. Self-esteem at endline is an indicator for above-median value on 5 items from the 10 point Rosenberg (1965) psychometric scale. Post-treatment, it is one item from the scale. Expected search time is measured as the number of months the candidate expects to search before getting a job divided by the number of job applications she plans to submit in the next 30 days. Expected o ers is measured as the number of o ers expected in the next 30 days, transformed by IHS. We do not report a multiple testing adjustment in this table because we measure three conceptually di erent outcomes: beliefs about skills, self-esteem, and beliefs about labor market prospects.

Table A.9: Treatment E ects on Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Employed	Month 1	Month 2	Hours o	Index	Earnings o	Hourly wage o	Written contract	Index
Public	0.052	0.036	0.058	0.201	0.138	0.338	0.197	0.020	0.106
	(0.011)	(0.011)	(0.014)	(0.052)	(0.025)	(0.074)	(0.040)	(0.010)	(0.028)
Private	0.011	0.028	0.008	0.066	0.050	0.162	0.095	0.017	0.065
	(0.012)	(0.013)	(0.015)	(0.048)	(0.028)	(0.078)	(0.046)	(0.009)	(0.030)
Placebo	0.020	-0.021	0.051	0.039	0.035	0.068	0.054	0.005	0.028
	(0.027)	(0.026)	(0.029)	(0.075)	(0.064)	(0.185)	(0.129)	(0.021)	(0.064)
q: public = 0	0.001	0.001	0.001	0.001		0.001	0.001	0.020	
q: private = 0	0.522	0.136	0.522	0.346		0.067	0.067	0.067	
q: placebo = 0	0.831	0.831	0.471	0.831		1.000	1.000	1.000	
q: public = placebo	0.196	0.098	0.485	0.098		0.665	0.665	0.665	
# observations	6605	6602	6605	6596	6607	6587	6572	6573	6607
# clusters	84	84	84	84	84	84	84	84	84

Coe cients are from regressing each outcome on a vector of treatment assignments, randomization block xed e ects, and prespeci ed baseline covariates (measured skills, self-reported skills, education, age, gender, employment, discount rate, risk aversion). Heteroskedasticity-robust standard errors shown in parentheses, clustering by treatment date. All outcomes use a 7-day recall period unless marked with z (30-day recall period) or y (since treatment). Outcomes marked with o use the inverse hyperbolic sine transformation. q-values control the false discovery rate within each family of outcomes.

E Certi cates Used in Skill-blind Treatment

Figure A.1: Sample Skill-Blind Certi cate

Notes: This gure shows an example of the certi cates given to candidates in the skill-blind treatment group. The certi cates contain the candidate's name and national identity number, and the logo of the World Bank and the implementing agency. Each work seeker received 20 of these certi cates and guidelines on how to request more certi cates.

F Audit Study

To identify the effect of information provision on the demand side, we are conducting an audit study (still in the field). We submit real workseekers' applications to entry-level job vacancies we sourced from a number of online job posting sites. We directly vary the information firms see about workseekers' skills, by randomizing whether the applications include information about workseekers' assessment results. This design allows identifying the effect of alleviating demand-side information frictions without conditioning on workseekers' decisions to share their assessment results with firms.

We implement the audit study in nine sequential rounds. Rounds 1-7 were already completed (Appendix Table A.10) and rounds 8-9 are still in the field. In each round, a subset of candidates who have completed the workseeker study endline is randomly selected and invited by text message to submit application materials to us, within 7 days, for an undisclosed job opportunity. We do not explicitly indicate our affiliation or a specific institution or organization for the job openings to avoid making participants more or less likely to apply.

Table A.10: Implementation Details of Audit Study Rounds 1 to 7

	Cumulative (1-7)	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7
Panel A: Search Intensity								
Candidates invited	1,860	204	378	270	234	234	270	270
Candidates responded	546	66	126	68	76	71	87	52
Panel B: Audit Study								
Vacancies	1,000	157	210	106	119	140	124	144
Applications attempted	3,999	628	840	424	475	560	496	576
Responses received	498	55	129	35	90	55	76	58

We send each individual a text message:Dear <name>, we have identified a job opportunity for you. We are a group of researchers trying to help young people find jobs. If you are interested, email your CV to <email address> or fax your CV to <fax number>. Find more info at <website>. Please send your CV within 7 days. A CV (curriculum vitae) in South Africa is generally understood to include all materials relevant to job applications. One additional reminder text message is sent to all candidates 1-3 days after this initial message. Once a candidate sends their application, they receive an automated acknowledgement.

Approximately 30 percent of the work seekers across all rounds contacted responded to our message within a week. Work seekers participating in the audit study are slightly selected sample. As observed in Appendix Table A.11, compared to work seekers in the supply-side study sample they are more likely to be male, and to have post-school qualifications. They perform worse in the numeracy and literacy tests, and slightly better in concept formation and grit. Importantly, they are equally likely to have worked in the past 7 days, and slightly more likely to have been assigned to the private treatment arm in the supply-side study.

We process the applications received and record information on when the application was re-

Table A.11: Comparison Between Audit and Workseekers Study Samples

	Audit study sample			Workseeker sample		
	Mean	Std Dev.	Obs	Mean	Std Dev.	Obs
Panel A: Characteristics of Applications Received from Workseekers						
Includes references or a reference letter	0.88	0.32	527	-	-	-
Includes a cover letter	0.14	0.34	528	-	-	-
Includes a copy of ID document	0.49	0.5	480	-	-	-
Includes information about matric	0.6	0.49	524	-	-	-
Includes references or a reference letter	0.88	0.32	527	-	-	-
Panel B: Characteristics of Workseekers						
Public treatment	0.31	0.46	543	0.33	0.47	6,891
Placebo treatment	0	0	543	0.037	0.19	6,891
Private treatment	0.36	0.48	543	0.31	0.46	6,891
Age	23.84	3.18	543	23.65	3.3	6,891
Male	0.48	0.5	543	0.38	0.49	6,891
Completed diploma or degree	0.19	0.39	543	0.17	0.37	6,891
Completed post-high school certi cate	0.24	0.43	543	0.21	0.41	6,891
Completed high school	0.56	0.5	543	0.61	0.49	6,891
Completed less than Grade 12	0.44	0.5	543	0.39	0.49	6,891
Numeracy assessment score (z score)	0.037	0.96	543	0.052	0.99	6,891
Literacy/communications assessment score (z score)	-0.0049	0.93	543	0.05	0.99	6,891
Concept formation assessment score (z score)	0.084	0.92	543	0.047	0.99	6,891
Grit assessment score (z score)	0.15	1	543	0.031	0.99	6,891
Worked in the last 7 days (endline)	0.39	0.49	543	0.4	0.49	5,836

ceived, where it was sent from, and what each individual application contains. Workseekers' response rates or search intensity are recorded as an observed measure of search behavior. This allows us to identify the e ect of workseeker-level treatment assignment on the decision to apply for jobs.

We identify entry-level job vacancies from a number of online job posting sites. Selected vacancies are suitable for entry-level workers, such that all candidates in our sample would be eligible to apply. We exclude jobs that look suspicious or are discriminatory, for example: jobs that ask for payments of any kind, or promise unrealistic salaries or bene ts, or discriminate based on appearance, race, or gender. The curated list rarely exceeds 200 vacancies per round. Appendix Table A.12 describes the sample of vacancies. Typical sectors include retail, sales, admin, restaurant, security, packing and warehouse related, and call centers.

For each participating work seeker who responded to our invitation, we prepare and submit applications to multiple job vacancies. We send each vacancy 4 job applications from di erent work seekers. We try to minimize the time spent between sourcing and sending job applications to increase the likelihood that vacancies are still open at the implementations point.[38] We generate between

[38] Given our implementation design, there may be up to a two week lag between the time we receive CVs and when we send applications on behalf of the candidates this is to allow for us to build and curate job vacancies, and to allow enough time for candidate submissions to accumulate. However, job vacancies may become lled during that wait period.

Table A.12: Vacancy-Level Attributes

	# Obs	Mean	Std Dev.
Panel A: Job Sector			
Sales	964	0.47	0.50
Admin	964	0.22	0.41
Service	964	0.17	0.38
Call centre	964	0.12	0.32
Industrial	964	0.09	0.29
Recruitment	964	0.08	0.27
Uncategorised	964	0.15	0.35
Panel B: Responses to Applications Submitted			
Responds to any application	964	0.14	0.34
Responds to all applications	964	0.07	0.25

6 and 10 applications per work seeker in each round completed to date, so that the total number of applications equals 4 times the number of jobs. We do not represent ourselves as the candidate. Instead, a generic message and subject line are written for each of the four email addresses. Subject line: Application for <vacancy> / Application for <candidate name>. Body: Please find attached the application for <vacancy> as recently advertised online. / Please find the application for <candidate name> > for <vacancy>, as recently advertised online .

We assign treatment status at the vacancy-application level. We employ a within-unit randomization design similar to Abel et al. (2019), with the difference that for each vacancy they select job seekers who have previous work experience in a related sector. We randomly assign the applications generated for each work seeker to treatment or control status. Treatment applications include a public report. Control applications include no report. In all other respects, treatment and control applications are identical. Importantly, the application treatment is independent of workseekers' treatment status in the workseekers' study and of their decision to include a report in the CV they submit to us. Further, we randomly assign each vacancy to "high" or "low" treatment saturation. High treatment saturation vacancies get a public report in 3 of the 4 applications submitted. Low treatment saturation vacancies get only 1 application with a public report attached.

We monitor and record responses for up to two weeks and inform candidates of any interview requests or job offers. We screen out responses that seem illegitimate or are identified as automated. Then we establish whether the response falls in one of the following categories: an acknowledgement of receipt , a request to send more information , an interview request , a request to visit the establishment in person , a job offer , a rejection , a scam , or whether the vacancy has closed. We construct outcome indicators for whether the application received any response (acknowledgement of receipt, rejection, request for more information, request to visit business, or interview/shortlisting), and whether the response was an interview invitation.

As shown in Appendix Table A.13, roughly half of a workseeker's applications are assigned to be control and half to the public report treatment. Half the applications from each workseeker are sent to high saturation vacancies and half to low saturation vacancies. Of all applications submitted,

Table A.13: Descriptive Statistics for Application-Level Attributes

	# Obs	Mean	Std Dev.
Had one report in a vacancy with one report	3,819	0.12	0.33
Had one report in a vacancy with three reports	3,819	0.37	0.48
Had no report in a vacancy with one report	3,708	0.37	0.48
Had no report in a vacancy with three reports	3,708	0.13	0.33
Any response received	3,579	0.14	0.35
Interview request received (uncon)	3,579	0.08	0.28
Acknowledgement received (uncon)	3,533	0.01	0.12
More information requested (uncon)	3,512	0.03	0.18
Scam, rejected, closed (uncon)	3,588	0.00	0.07
Interview request received (con)	507	0.59	0.49
Acknowledgement received (con)	507	0.10	0.29
More information requested (con)	507	0.24	0.43
Scam, rejected, closed (con)	507	0.03	0.18

only a small fraction (14 percent) receives any type of response, and slightly more than half of those receiving a response obtain an interview request (8 percent of the full sample of applications).

The audit study is designed to directly identify the average e ect of alleviating pure demand-side frictions. Estimates exploit within-applicant randomization, comparing responses to applications submitted with and without a public report. Thus, for the sample of audit study participants, the average treatment of providing more information to rms is identi ed without adjusting for selection. Although our study is not designed to identify spillover e ects, our design allow us to provides some insight by comparing responses to applications submitted with a public report, across vacancies that receive one and three applications with a public report attached.

The audit study design described above does have some limitations. It only identi es an e ect on demand-side decisions and only at one stage of the process. It does not allow us to reliably observe supply-side responses, such as decisions to accept or reject interview invitations. Also, it does not allow us to reliably observe equilibrium outcomes, such as job o ers and acceptances after interviews. We thus view the audit and workseekers' studies as complements, that jointly identify parameters of the job search and hiring environment that neither individual study could identify.

G Screenshots of Platform Used in Firm-Facing Experiment

Figure A.2: Screenshots of Login Page and Filtering Page