Job Search and Hiring with Two-sided Information Frictions*

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

October 3, 2019

Abstract

Firms make hiring decisions and workseekers make job search decisions using potentially noisy signals of workseekers' productivity in different jobs. Noise can distort job search and hiring decisions and lead to lower total employment and earnings. We study the labor market effects of giving better information about workseekers to firms and/or workseekers. Assessing workseekers' skills in multiple domains, certifying their assessment results, 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, showing that both workseekers and firms face information frictions. These findings demonstrate quantitatively important information frictions on both sides of the labor market that can be alleviated by assessment and certification.

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

*We are grateful for helpful comments from Peter Arcidiacono, Rulof Burger, Stefano Caria, Jishnu Das, Erica Field, Rachel Heath, Lisa Kahn, Melanie Morten, Duncan Thomas, Chris Woodruff and seminar participants at the NBER Summer Institute, Duke, Georgia, Notre Dame, NYU, Oxford, UIUC, and the World Bank. We appreciate the exceptional research assistance from Wim Louw, Wendy Trott, Shelby Carvalho, Allegra Cockburn, Caitlan Russell, 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://doi.org/10.1257/rct.1631-8.0. 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

Firms make hiring decisions and workseekers make job search decisions using potentially noisy signals of workseekers' productivity. Firms decide if and whom to hire, and what wage to offer, based on signals of productivity like education and experience. Workseekers decide how much to search and how to direct that search based on feedback on their performance in education and in past work. If these signals are weak predictors of performance in specific jobs, then one or both sides of the market face information frictions. Information frictions for firms can reduce the expected marginal revenue product from new hires and hence reduce wages. Information frictions can also reduce firm-level employment when firms face binding minimum wages or firing costs. Information frictions for workseekers can distort job search decisions. These search distortions can also change individual-level employment and earnings. For example, overconfident workseekers might reject job offers with low wages and hence remain unemployed, while underconfident 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 qualifications are poor proxies for skill.

We use a series of field experiments to identify how revealing information about workseekers' types changes their beliefs about themselves 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 – where information frictions are likely to bind: formal education is weakly correlated with measured skills, minimum wages bind, and firms face high separation costs. We study a population where information frictions are likely to bind: 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 use psychometric assessments to measure their skills in six domains: communication, concept formation, focus, grit, numeracy, and planning.

In our first experiment, we randomly select some workseekers for a 'public' certification intervention. This gives them their assessment results and allows them to share their results with firms when they apply for jobs. Relative to a control group who are assessed but receive no intervention, certification increases workseekers' employment rate from 31 to 36% after three to five months. Certification also increases weekly earnings by 34%, due to both higher employment and higher earnings conditional on employment. These findings suggest that the information provided by certification allows firms to hire better matches and potentially make more hires. Certification is more effective for workseekers with fewer alternative ways to signal their type: those with less work experience, less education, and worse latent employment prospects. The effects of certification do not vary substantially by workseeker type, potentially because our assessments mainly differentiate workseekers horizontally rather than vertically, showing areas of relative strength and weakness. Consistent with this interpretation, we show there is substantial heterogeneity in firms' relative demand for different workseeker types in a small incentivized resume-ranking experiment.

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, certification may help workseekers to learn about their type and hence better target their job search. But it may instead help firms learn about workseekers' types and hence make more, better-targeted job offers. We use additional outcome measures and additional experiments to separate these hypotheses.

First, we examine certification effects on workseekers' behavior. We find that certified workseekers use their certification in job applications, have more accurate beliefs about their types, target their search toward jobs that match their types, and expect better labor market outcomes. This shows that certification changes workseekers' behavior. But this change might for two reasons: workseekers may learn about their own types from certification, or workseekers may expect firms will learn about their types from certification.

Second, we conduct an intervention that shifts workseekers' information, but not firms' information. We randomly select some workseekers for a 'private' certification intervention. This gives them their assessment results in a way that is difficult to share with firms.[1] While public certification is conceptually similar to signals like formal education qualifications, private certification is conceptually similar to private information like skills assessments offered as part of job search counselling at labor centers. Private and public certification have similar effects on workseekers' beliefs about their type and their job search targeting. But private certification has smaller effects on employment and earnings than public certification. This shows that workseekers have limited information about their types and change behavior in response to information. However, giving them information without help signalling it to firms has limited effects on their labor market outcomes.

Third, we conduct an audit study that shifts firms' information, but not workseekers' information. We submit applications to real job vacancies using real resumes from workseekers in our sample. We submit multiple applications per resume and per vacancy, randomizing at the resume level whether applications include public certifications and randomizing at the vacancy level how many applications include public certifications. Applications including reports get more callbacks and vacancies getting more applications with reports make more callbacks. But the application-level gain from having a report is smaller when competing against other applications with reports. This shows that firms have limited information, that providing more information increases one measure

[1] Workseekers in the public certification intervention receive 20 printed and an emailed certificates with their assessment results. The certificates show the workseekers' name and national identity number and are branded by the World Bank and the prominent social enterprise that conducts the assessments. Workseekers in the private certification receive 1 printed certificate with their assessment results but without their name, their national identity number, or any branding. Both groups of workseekers receive a group session with an industrial psychologist who explains the assessment results. Control group workseekers do not learn their assessment results.

of labor demand, and that the effect of information is partly attenuated at scale. We also show that firms value more information using a small incentivized choice experiment, which measures firms' willing to pay for access to a database of certified workseekers.

Taken together, these results show 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 run one more small experiment to confirm that certification works by providing information on workseeker types. We give some workseekers certificates describing the assessments but withholding their assessment results; this does not increase employment..

We build on a growing literature on information frictions in labor markets. Several panel studies show employer learning over time about their workers' productivity, after observing noisy proxies at the time of hiring (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). These information frictions can have macroeconomic implications, increasing frictional unemployment (Donovan et al., 2018).

We make two major 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. The incidence 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 give firms better information. This might be performed by in-house assessment teams or specialized third-party assessment firms. 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 never appear in the pool they assess. These factors also influence which side of the market, if either, will be willing to pay for access to better information.

Second, we provide the first evidence that alleviating information frictions can substantially increase employment for inexperienced workseekers. We interpret the employment effects as evidence that information frictions reduce the risk-adjusted value of some hires below a minimum or reservation wage, consistent with the model in Pallais (2014). We study a setting with minimum wages, high reservation wages due to transport costs, and firing regulations that increase the downside

[2] A related literature tests if workers with different education qualifications but similar measured skills have different labor market outcomes (Alfonsi et al., 2017; Jepsen et al., 2016; 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. Autor and Scarborough (2008) study a large firm adopting new screening technology that generated only firm-side information.

risk of bad hires. Prior experiments studying information frictions for inexperienced workseekers have worked in lightly-regulated labor market and found no employment effects.[4] The employer learning literature using US panel data looks only at earnings trajectories for employed workers. The employment margin for inexperienced workseekers is particularly important in the context of information frictions. Without work experience, workseekers cannot start revealing their types to themselves or to firms through employer learning. This may lead to privately optimal but socially suboptimal firm preferences for experienced workers: firms who hire inexperienced workseekers face the downside risk of bad matches and may lose good matches to other firms once they acquire experience (Acemoglu and Pischke, 1999; Kahn, 2013). The certification we study offers a relatively cheap way to improve transitions into first employment in the many regulated labor markets facing high youth unemployment.

Our work also relates to a broader literature on the role of information in job search and hiring. Firms may use referrals through networks, reference letters or performance evaluations from past jobs, or internships and apprenticeships to learn about workseekers' types. Each of these strategies has some limitations. Referees may be motivated by family or social ties, limiting productivity gains from referrals and entrenching cross-group inequality in the labor market.[5] Reference letters or performance evaluations are easier for experienced workseekers to obtain than inexperienced or first-time workseekers.[6] Apprenticeships and internships can impose high time and monetary costs for firms and/or workseekers.[7] The certifications we study provide information quickly and cheaply, require no work experience, and can be distributed to any prospective employer without requiring network connections. We do not claim that assessment-based certifications are a perfect solution or eliminate the need for other screening tools. We find some suggestive evidence that certifications make referrals more accurate or credible: job offers after referrals account for the largest share of increased employment amongst certified workseekers. The certifications we study also help workseekers learn about their types and better target job search, which complements research showing that workseekers change search behavior in response to new information about labor market opportunities (Ahn et al., 2019; Altmann et al., 2018; Belot et al., 2018).

Our work is also relevant to the large literature on active labor market programs (ALMPs). These programs are common in developed and developing countries, despite mixed evidence on

[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 earnings.

[5] Beaman and Magruder (2012), Burks et al. (2015), Heath (2018), and Pallais and Sands (2016) show that hiring through referrals can lead to higher worker productivity. Beaman and Magruder (2012), Beaman et al. (2018), and Witte (2019) show that reference-based hiring does not necessarily identify the most productive candidates and can contribute to inequality. Ioannides and Loury (2004) review the broader literature on referrals in the labor market.

[6] See Pallais (2014) and Abel et al. (2019) for evidence that references from past employers can increase employment, respectively for gig workers on an online platform and for women but not men in the South African labor market.

[7] Hardy and McCasland (2017) show that small firms can use apprenticeships to screen prospective workers, but that these apprenticeships can be expensive.

their effectiveness (Card et al., 2015; McKenzie, 2017). The certifications we study can be cheaply and relatively easily included in ALMPs, which might enhance their effectiveness. This complements other work showing that tweaking ALMPs to increase their information content can enhance their effectiveness (Abel et al., 2019; Belot et al., 2018; Wheeler et al., 2019).

In Section 2 of the paper, we describe the economic environment: a sketch of our conceptual framework, the labor market and sample we study, and the psychometric assessments. In Section 3, we describe the skill certification experiment and report treatment effects on labor market outcomes. In Section 4, we explain the treatment effects on labor market outcomes using our additional experiments. We conclude in Section 5 and briefly discuss questions around scale effects, general equilibrium, and markets for assessment-based certification.

2 Economic Environment

2.1 Conceptual Framework

We briefly sketch a static conceptual framework here and discuss details and extensions in Appendix A. We consider an economy where workseekers have heterogeneous types T_i but are otherwise homogeneous. Workseekers allocate time between formal job search and other activities (e.g. leisure) based on their perceived probability of a job offer conditional on search and expected wage conditional on a job offer. Firms have heterogeneous production technology, such that the marginal revenue product of a worker with type T_i differs across firms. The probability that a firm meets a workseeker is an increasing function of the workseekers' search effort. Each firm hires all workseekers that they meet whose type is in a target range and pays them their marginal revenue product.

Information frictions can enter this environment in three main ways. First, workseekers may observe their types with error while firms observe their true types. This might occur if schools give students noisy feedback on their performance or workseekers see noisy measures of their performance in past jobs. Workseekers will make job search decisions based on their perceived type. The direction of the effects on search effort and total wages depend on the specific structure of the model. If the return to search effort is independent of type, search decisions are not distorted. If the return to search effort varies by type, then some workseekers will search too much and others too little. Firms will not meet some workseekers they would otherwise hire, potentially reducing employment and total wages.

Second, workseekers may observe their true types while firms observe workseekers' types with error. This might occur if attributes observable to firms, like educational qualifications or past work experience, are noisy proxies for productivity. Workseekers will search 'correctly' but firms know they may hire some workseekers with types outside their target range. If marginal revenue product is linear in type and firms are risk-neutral, then firms will hire the same number of workers and pay the same total wage bill, but will pay some workers the wrong amount. If marginal revenue product is concave in type or firms face uninsurable risks from bad hires, then total wages may fall. Risks such as bad workers damaging capital equipment, offending customers, or needing severance pay when fired might not be insurable. Information frictions may drive some workseekers' risk-adjusted value to some firms below minimum or reservation wages, lowering total employment.

This simple framework illustrates that information frictions on either side of the market can lower total employment and total wages. Information frictions facing workseekers also distort search behavior, while information frictions facing firms do not. This generates two testable predictions, which we assess in Sections 3.3 and 4. Information frictions should distort employment more for workseekers with noisier signals of their type, a prediction we assess in Section 3.4. The incidence of frictions across workseekers of different types depends on the structure of the model.

Third, information frictions might affect both workseekers and firms. Depending on the structure of the model, frictions on both sides of the market might interact to generate larger distortions or partly offset each other. We do not explore this in detail, because our experiments are designed to test for the existence of information frictions facing each of workseekers and firms, not to identify their interaction.

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 many labor markets, particularly 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.[8] 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.[9]

Second, South African firms report they find it difficult to screen workers or struggle to find workers with suitable skills.[10] Grades and grade progression in schools serving poor communities

[8]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.

[9]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%).

[10]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,

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.[11] 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, 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, 2009). 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.[12] 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).

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).

[11] 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).

[12] 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).

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.[13]

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.[14] 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.[15] Our sample consists of all candidates who arrive at Harambee for the second of these two testing days, on 84 operational days.[16]

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

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

[14]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.

[15]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.

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

3-4 months after treatment.[17] The phone survey response rate is 96%, balanced across treatment groups (Appendix Table A.8), and related to few baseline covariates (Appendix Table A.9). 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.4 and A.5. 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.

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 difficult transitions into the labour market, despite having higher levels of education, so may be more likely to apply to Harambee.[18]

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).[19] 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

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

[18]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).

[19]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).

work in urban areas, an average of 560 ZAR (40 USD) per week at baseline.[20] 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 E.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 firms.[21] Among those in family firms or self-employment, 87% were in firms with four or fewer 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.[22] 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.[23] Thus certificates to attach to a written application are a useful way of communicating information about skills to employers.

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

[21] The breakdowns of formal workers by firm 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%.

[22] 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.

[23] 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.

2.5 Assessments

We conduct six assessments with workseekers: communication, concept formation, numeracy, focus, grit, and planning. Appendix B 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 described soft skills of these types as valuable.

Firms appear to value information about workseekers' assessment outcomes. First, they already use them. Harambee has used different combinations of the numeracy, communication, and concept formation assessments 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 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 school-leaving certificate with test scores. Second, we show in Section 4.4 that other firms have positive willingness-to-pay for better information on workseeker assessment results in a small incentivized choice experiments. There is substantial variation in firms' relative ranking of performance on different assessments, showing heterogeneous demand for different worker types.

In Appendix D, we propose a nonparametric test to establish if assessment results 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.

Assessments are conducted over two days. Each assessment session is led by two to three industrial psychologists, who manage a team of facilitators. Assessments are conducted in English and are self-administered on desktop computers.[24] Table A.4 shows standardized scores on all six assessments. There is a fairly even spread of candidates over the distribution and little evidence of ceiling effects.

The certificates display only ordinal results, showing the tercile in which the candidate placed on each assessment.[25] There was no way for candidates to access their cardinal result or for control

[24] Before assessments, candidates do practice computer exercises. However, some candidates have poor computer skills, so assessment results will be linked to candidates' computer proficiency. To minimize this link, no assessments require fast responses and the assessments are designed to be completable within the available time limit. Before starting assessments, candidates consent their assessment results being shared with Harambee, the research team, and external firms.

[25] The terciles are based on assessment results from candidates assessed before the study started: 5,000 workseekers for communication, numeracy and concept formation test, and 500 workseekers for focus, grit, and planning.

group workseekers to access their results.

Scores are weakly correlated across assessments, so the assessments mainly identify horizontally differentiated workseeker types, rather than vertically differentiated types. Few candidates have all bottom terciles (0.7%) or all top terciles (2.3%). Most candidates have at least one bottom tercile (75.7%) and at least one top tercile (88.1%); most cadidates have at least one top and at least one bottom tercile (63,9%). The first principal component of the six cardinal assessment results explains only 37% of their joint variation.

Workseekers have inaccurate beliefs about their own types. We ask workseekers in which tercile they believe they ranked after taking the assessments but before treated workseekers learn their results. 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 Certification

3.1 Intervention

We design a certification intervention which gives candidates information about their assessment results and allows them to share the results with 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 type of candidates who do not share certificates. We certify several thousand workseekers in a metropolitan area of almost ten million people. To the best of our knowledge, there is no other provider of psychometric assessment certifications in South Africa.

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 assessments and directs the reader to https://www.assessmentreport.info/ for more information on Harambee and the assessments. The certificate shows the tercile in which the candidate ranked on each assessment, compared to a benchmark group of other candidates assessed by Harambee.[26] To link candidates with certificates, each certificate shows the candidate's name and unique national identity number.[27] To provide credibility to the assessments and results,

[26]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.

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

the certificate is branded with the World Bank and the Harambee Youth Employment Accelerator logos.[28]

In the briefing, the psychologist explains how to interpret each of the assessments 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 certification and a control group. We randomize treatment by day to reduce risks of spillovers between treated and control workseekers.[29] Treated workseekers receive the certification intervention. Control group workseekers receive no information about their assessment results and no assistance sharing results with firms.

We estimate treatment effects using models of the form

$$Y_{id} = T_d \cdot \beta + \mathbf{X}_{id} \cdot \Gamma + S_d + \epsilon_{id}, \tag{1}$$

where Y_{id} is the outcome for workseeker i assessed on date d, T_d is the treatment assignment, $\mathbf{X}_{id}$ 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, 5, 6, 4, A.11, and A.12 and discuss 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

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

[29]Randomization is sequential and stratified, with days randomized within blocks of 6-10 upcoming days. Table A.4 shows that the randomization generates balanced treatment assignments. There are respectively 2,247 and 2,274 workseekers in the certification and control group, spread over 54 days. The treatment occurs on the second day of assessments when workseekers have completed all assessments.

Figure 1: Sample Public Certificate

THE WORLD BANK

harambee
YOUTH EMPLOYMENT ACCELERATOR
WORK FOR WORK

REPORT ON CANDIDATE COMPETENCIES

name.. surname..
ID No. id..

This report provides information on assessments conducted by Harambee Youth Employment Accelerator (harambee.co.za), a South African organisation that connects employers looking for entry-level talent to young, high-potential work-seekers with a matric or equivalent. Harambee has conducted more than 1 million assessments and placed candidates with over 250 top companies in retail, hospitality, financial services and other sectors. Assessments are designed by psychologists and predict candidates' productivity and success in the workplace. This report was designed and funded in collaboration with the World Bank. You can find more information about this report, the assessments and contact details at www.assessmentreport.info. «name» was assessed at Harambee on 13 September, 2016.

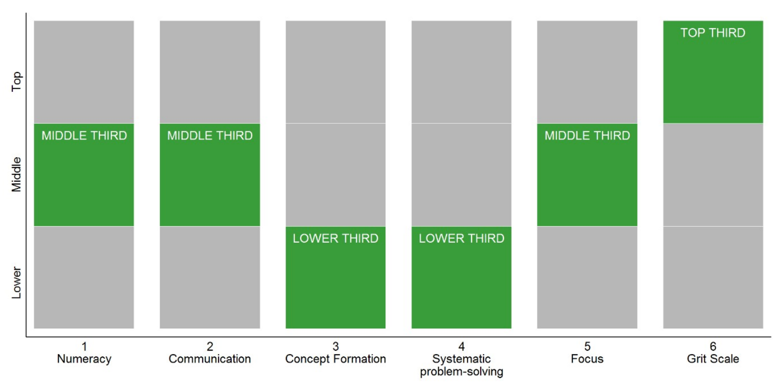
«**name**» completed assessments on English Communication (listening, reading, comprehension), Numeracy, and Concept Formation:

1. The Numeracy tests measure candidates' ability to apply numerical concepts at a National Qualifications Framework (NQF) level, such as working with fractions, ratios, money, percentages and units, and performing calculations with time and area. This score is an average of two numeracy tests the candidate completed.
2. The Communication test measures a candidate's grasp of the English language through listening, reading and comprehension. It assesses at an NQF level, for example measuring the ability to recognise and recall literal and non-literal text.
3. The Concept Formation Test is a non-verbal measure that evaluates candidates' ability to understand and solve problems. Those with high scores are generally able to solve complex problems, while lower scores indicate an ability to solve less complex problems.

«**name**» also completed tasks and questionnaires to assess their soft skills:

4. The Planning Ability Test measures how candidates plan their actions in multi-step problems. Candidates with high scores generally plan one or more steps ahead in solving complex problems.
5. The Focus Test assesses a candidate's ability to distinguish relevant from irrelevant information in potentially confusing environments. Candidates with high scores are generally able to focus on tasks in distracting surroundings, while candidates with lower scores are more easily distracted by irrelevant information.
6. The Grit Scale measures whether candidates show determination when working on challenging problems. Those with high scores generally spend more time working on challenging problems, while those with low scores choose to pursue different problems.

«name»'s results have been compared to a large benchmark group of young (age 18-34) South Africans assessed by Harambee. All candidates have a matric certificate and are from socially disadvantaged backgrounds. The benchmark group is 5,000 for cognitive skills and 400 for soft skills.
«name» scored in the «tercile_num» THIRD of candidates assessed by Harambee for Numeracy, «tercile_lit» THIRD for Communication, «tercile_cft» THIRD for Concept Formation, «tercile_tol» THIRD for Planning Ability, «tercile_troop» THIRD for Focus and «tercile_grit» THIRD for the Grit Scale.



DISCLAIMER: This is a confidential assessment report for use by the person specified above. The information in the report should only be disclosed on a "need to know basis" with the prior understanding of the candidate. Assessment results are not infallible and may not be entirely accurate. Best practice indicates that any organisation's career management decisions should depend on factors in addition to these assessment results. Harambee cannot accept responsibility for decisions made based on the information contained in this report and cannot be held liable for the consequences of those decisions.

Note: This figure shows an example of the certificates given to candidates in the 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.

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 Certification Improves Average Labor Market Outcomes

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.[30] 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 E.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).[31]

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. Certification has no effect on the rate of written permanent contracts, but these are rare in our sample, so this may be a difficult margin 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. 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 (Table A.10).

We can decompose the treatment effects on observed job attributes into extensive margin effects – explained by treatment effects on employment – and intensive margin effects – explained by

[30] 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.

[31] 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 Effects on Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)
Panel A: Employment Status Measures					
	Employed	Month 1	Month 2	Hours ≀	Index
Treatment	0.052***	0.036***	0.058***	0.201***	0.138***
	(0.011)	(0.011)	(0.014)	(0.052)	(0.025)
q: Treatment effect = 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 ≀	Hourly wage ≀	Written contract	Index	
Treatment	0.338***	0.197***	0.020*	0.106***	
	(0.074)	(0.040)	(0.010)	(0.028)	
q: Treatment effect = 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	

Note: Coefficients are from regressing each outcome on a vector of treatment assignments, randomization block fixed effects, and prespecified baseline covariates (assessment results, self-reported assessment results, 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 ‡ (30-day recall period) or † (since treatment). Outcomes marked with ≀ use the inverse hyperbolic sine transformation.

treatment effects on earnings for employed workers. This distinction is important. If certification allows workseekers to better target search or firms to hire better-matched workers, then mean match quality should be higher in the treatment group. We adapt a decomposition proposed by Attanasio et al. (2011). Intuitively, the extensive margin effect on earnings is simply the average treatment effect on employment multiplied by the mean earnings for employed control group members. The intensive margin effect on earnings is the average treatment effect on earnings minus the extensive margin effect. The same argument applies to hours and contract type but is not meaningful for hourly wages, which are already an intensive-margin effect. See Appendix C for more detailed derivation.

Job attributes, except for earnings, shift mainly at the extensive margin (Table 2). The hours and contract type effects are explained entirely by the extensive margin. In contrast, the extensive and intensive margin effects for earnings are respectively 27 and 7%, both significantly larger than zero. The 7% intensive margin effect and 36% employment rate in the certification group means that each employed candidate in the certification group earns 20% more than they would without certification. Hence, certification both increases employment and allows candidates to earn more when employed. This is consistent with an improvement in match quality, if proxied by hourly

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

	(1) Earnings $\wr$	(2) Hours $\wr$	(3) Written contract
Total effect	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 effect conditional on employment	0.195*	0.036	-0.001
	(0.113)	(0.058)	(0.024)
Control group mean	5.177	3.624	0.392

Note: This table reports decompositions of treatment effects on job characteristics into extensive and intensive margins. The extensive margins are the treatment effects on job characteristics due to the treatment effect on employment, evaluated at the mean job characteristics for the control group. The intensive margins are the residual treatment effects on job characteristics, which must be due to changes in job characteristics for the employed candidate in the treatment group. The conditional effect 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 $\wr$ use the inverse hyperbolic sine transformation.

wages or by earnings conditional on employment.

3.4 Certification Is Most Effective When Worseekers Lack Alternative Signals

Certification is likely to be most effective when workseekers lack alternative signals of their type. These signals might include past work experience and post-secondary education, which allow them to learn about their own productivity in specific tasks and signal this productivity to firms. We test this idea by augmenting equation (1) to include interactions between treatment and proxies for alternative signals. Treatment effects are 2.7 percentage points smaller for candidates with post-secondary education and 4.3 percentage points smaller for candidates with prior work experience, although these differences are not statistically significant (Table 3). We also estimate the latent probability of being employed at endline as a single summary measure of candidates' scope to get a job without certification.[32] Candidates with above-median latent probabilities of employment have 6.9 percentage point smaller treatment effects than candidates with below-median latent probabilities. These results show that certification is most effective for workseekers facing the most serious information frictions.

In contrast, we do not find clear heterogeneity in certification effects by assessment results.

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

Table 3: Treatment Effects on Labor Market Outcomes by Access to Alternative Signals

	(1)	(2)	(3)
Outcome	Employed	Employed	Employed
Treatment	0.052***	0.052***	0.051***
	(0.0.11)	(0.012)	(0.012)
× post-secondary education	-0.027		
	(0.028)		
× employed at baseline		-0.043	
		(0.028)	
× $\hat{\Pr}$(employed at endline|X)			-0.069**
			(0.028)
Mean outcome	0.309	0.309	0.309
# observations	6605	6605	6605
# clusters	84	84	84

Coefficients are from regressing each outcome on a vector of treatment assignments, displayed interaction terms, randomization block fixed effects, and prespecified baseline covariates (assessment results, self-reported assessment results, 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 alternative signal measures are demeaned before being interacted with treatment, so the coefficient on the treatment indicator equals the average treatment effect. $\hat{\Pr}$(employed at endline|X) is estimated by regressing endline employment status on the baseline covariates listed above, using only control group data, and predicting outcomes for both control and certification groups. Prediction for control group observations uses leave-one-out-estimation to avoid overfitting. All predicted values are strictly between zero and one.

We test for heterogeneity by assessment results in a variety of ways: we construct several single indices of assessment performance (e.g. using principal components) and interact these with treatment assignment, and we develop a new nonparametric test. We describe these tests and results in Appendix **??**. In brief, we find no robust evidence that treatment effects vary systematically with assessment results. This is consistent with the fact that assessment results mainly differentiate candidates horizontally, rather than vertically. Most candidates perform well in some assessments and poorly in others, while than 4% of candidates perform well or poorly in all six assessments (Section 2.5). Treatment effects may not vary by assessment results because candidates with different assessment results are valued by different types of firms.

We test the idea of heterogeneous demand for different worker types using an incentivized choice experiment with 69 firms. We ask firms to rank applications from seven hypothetical candidates and tell them we will use their ranking to match them with workseekers from our sample, following Kessler et al. (2019). Applications randomly vary in the assessment results they show: all middle terciles, or five middle terciles and one top tercile in a randomly chosen assessment. There is substantial variation in firms' ranking of assessments relative to each other. Focus is most popular, ranked highest by 32% of firms. But each of the other five assessments is ranked highest by 6 to 19% of firms. These results show that firms have heterogeneous demand for workseeker types, although the sample size is limited. We describe the firm sample and recruitment process in more detail in

Section 4.4.

4 Separating Workseeker- and Firm-side Information Frictions

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 types, while firms observe types 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' types perfectly, while candidates observe their types 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, certification may not provide useful information about candidates' types 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 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 types. 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 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 candidates whom firms misclassify and overpay. If only workseekers face information frictions, then firms can pay some workers below their marginal revenue product.

The incidence of welfare costs and optimal policy design depend on the 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.

Table 4: Treatment Effects on Self-Beliefs

	(1) % terciles correct	(2) Targeted search	(3) Expected search time	(4) Expected offers
Public	0.158***	0.052***	-0.037***	0.106***
	(0.008)	(0.010)	(0.013)	(0.019)
Private	0.123***	0.047***	-0.024	0.053**
	(0.008)	(0.010)	(0.015)	(0.023)
p: public = private	0.000	0.698	0.290	0.025
Mean outcome	0.389	0.155	0.408	1.807
# observations	6607	6609	6332	6531
# clusters	84	84	84	84

Coefficients are from regressing each outcome on a vector of treatment assignments, randomization block fixed effects, and prespecified baseline covariates (assessment results, self-reported assessment results, 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. % terciles correct is the fraction of assessments where the candidate correctly reports the tercile in which they scored. Targeted search is an indicator equal to one if the candidate reports mainly applying for jobs that most value the skill in which the candidate believes they scored highest. 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 offers is measured as the number of offers expected in the next 30 days, transformed by IHS. We do not adjust these results for multiple testing because we measure four conceptually different outcomes.

4.1 Certification Changes Job Search and Beliefs

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 certification has a large effect on candidates' beliefs about their types and their job search (Table 4). We ask candidates at baseline and endline if they think they were in the top, middle, or bottom third of candidates on each assessment. Certification increases the fraction of assessments where candidates' self-assessments match their measured results from 0.39 to 0.45.[33] We also measure candidates' search targeting: we ask candidates if the types of jobs they are applying for most value communication, concept formation, or numeracy. Certification increases the fraction of candidates searching for jobs that most value the assessment in which they scored highest from 0.16 to 0.21. Certification also increases two measures of perceived search effectiveness: candidates believe they will get jobs in 9% less time (counterfactual mean 0.4 months) and will get 11% more offers in the next month.

[33] The treatment effect on accuracy of beliefs about assessment results is already visible in the a text message survey conducted 2-3 days after treatment (Table A.11). Treatment effects on a short self-esteem scale are precise zeroes in the endline phone survey and the text message survey. This suggests certification adjusted targeted beliefs about types and employment prospects rather than general self-evaluation.

Candidates use certificates heavily (Table 5 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. Candidates with fewer alternative signals of their type (no post-secondary education, less work experience) do not use certificates more. Combined with their larger employment effects, this suggests the certificates add more value per application for these candidates.

However, treated candidates do not search more or more effectively in the week or month before the endline survey. 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 5 panel B). Similarly, certification does not increase the number of responses or offers received from employers in the preceding month (Table 5 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 by 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.[34]

Treatment effects on on-the-job search are also consistent with this explanation. 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, 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

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

Table 5: Treatment Effects on Job Search

	(1)	(2)	(3)	(4)	
Panel A: Certificate Use Measures					
	Any use †	Applications ≀†	Interviews †	Offers †	
Treatment	0.699***	1.683***	0.432***	0.112***	
	(0.013)	(0.040)	(0.023)	(0.011)	
q: Treatment effect = 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 Effort Measures					
	Any search	Apps †≀	Search hours ≀	Search cost ≀	Index
Treatment	-0.020	0.018	-0.034	-0.092	-0.013
	(0.014)	(0.042)	(0.048)	(0.081)	(0.032)
q: Treatment effect = 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 Effectiveness Measures					
	Responses ‡	Offers ‡	Index		
Treatment	0.049	0.007	0.025		
	(0.047)	(0.017)	(0.029)		
q: Treatment effect = 0	1.000	1.000			
Mean outcome	0.772	0.182	0.000		
# observations	6591	6590	6591		
# clusters	84	84	84		

Note: Coefficients are from regressing each outcome on a vector of treatment assignments, randomization block fixed effects, and prespecified baseline covariates (assessment results, self-reported assessment results, 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 ‡ (30-day recall period) or † (since treatment). Outcomes marked with ≀ use the inverse hyperbolic sine transformation. We do not construct an inverse-covariance weighted average of the certificate use measures because these do not vary in the control group, making it impossible to estimate the control group covariance matrix.

suggests that the certificate might enhance the effectiveness of referrals, by helping network links to target referrals or making their referrals more credible to employers.

Taken together, these results are consistent with candidates updating their beliefs about their types, 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.

4.2 Workseekers Face Information Frictions

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 same experiment. These candidates receive one copy of an unbranded, anonymous certificate with the 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 public 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 call this the 'private' treatment intervention, to distinguish it from the 'public' treatment arm described above.

We interpret the private treatment as primarily providing information to the candidates about their own types. Candidates can share the private certificate with firms but this is less likely to change firms' decisions than the public 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 types (Table A.11). These effects are very slightly smaller than the corresponding public treatment effects on beliefs at the endline but identical in the text message survey conducted soon after treatment. The private and public treatments have the same effect on search targeting. The private treatment reduces the expected search time required to get a job and increased the expected number of job offers in the next month. Both effects are smaller than in the public treatment, though only the latter difference is significant. This shows that giving workseekers better information updates their beliefs about their types, helps them target their job search, and slightly improves their expected search outcomes. But the small effect on expected search outcomes suggests that workseekers view the public certification as more valuable in the labor market.

Candidates in the private treatment group do share their certificates with employers, but substantially less often than candidates in the public treatment group (Table 6 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.

Figure 2: Sample Private Certificate

REPORT ON CANDIDATE COMPETENCIES
-Personal Copy-

This report contains results from the assessments you took at Harambee in Phase 1 and Phase 2. These results can help you learn about some of your strengths and weaknesses and inform your job search.

You completed assessments on English Communication (listening, reading and comprehension) and Numeracy today in Phase 2. In Phase 1, you completed a Concept Formation assessment which asked you to identify patterns.

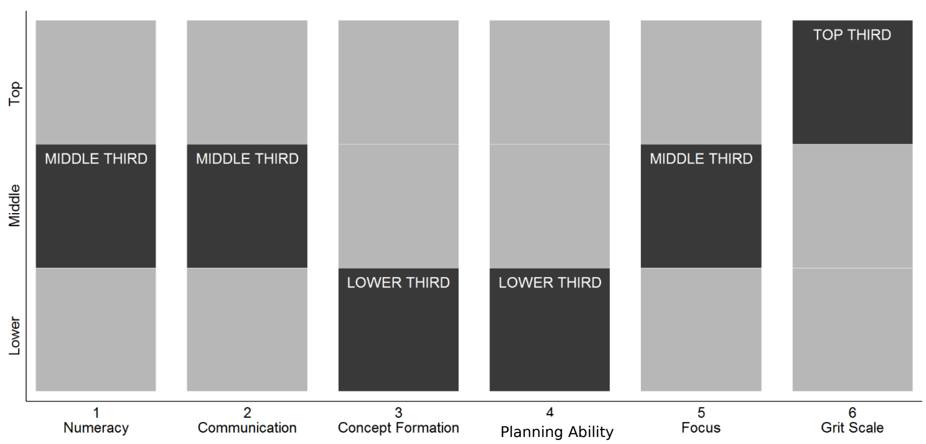
1. The Numeracy tests measure various maths abilities. Your score is the average of the two maths tests you did today at Harambee.
2. The Communication test measures English language ability through listening, reading and comprehension.
3. The Concept Formation test measures the ability to understand and solve problems. Candidates with high scores can generally solve complex problems, while lower scores show an ability to solve less complex problems.

You also did some games and questionnaires to measure your soft skills:

4. The Planning Ability Test measures how you plan your actions in multi-step problems. Candidates with high scores generally plan one or more steps ahead in solving complex problems.
5. The Focus Test looks at your ability to pick out which information is important in confusing environments. Candidates with high scores are able to focus on tasks in distracting situations.
6. The Grit Scale measures candidates' determination when working on difficult problems. Candidates with high scores spend more time working on the problems rather than choosing to pursue different problems.

Your results have been compared to a large group of young South African job seekers who have a matric certificate, are from socially disadvantaged backgrounds and have been assessed by Harambee.

You scored in the MIDDLE THIRD of candidates assessed by Harambee for Numeracy, MIDDLE THIRD for Communication, LOWER THIRD for Concept Formation, LOWER THIRD for Planning Ability, MIDDLE THIRD for Focus and TOP THIRD for the Grit Scale.



DISCLAIMER
Please note that this is a confidential assessment report and is intended for use by the person specified above. Assessment results are not infallible and may not be entirely accurate.

Note: This figure shows an example of the certificates given to candidates in the private treatment arm. The certificates contain the candidate's assessment results but no identifying information and no branding. Each candidate received one copy of this certificate

Like the public treatment, the private treatment 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 substantially significantly smaller than the corresponding public treatment effects. We cannot reject that the intensive margin effects of the private and public treatments are equal, using the decomposition from Appendix C.

These findings are consistent with multiple possible interpretations. Private certification may convey information only to workseekers and hence change beliefs about skills and returns to search, job search, and labor market outcomes. As the private certification effects on labor market outcomes 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 types. The fact that workseekers' beliefs about types change just as much in response to information, even if they are less able to reveal it to firms, suggest they face at least some information frictions.

We cannot completely rule out an alternative explanation, that there is some firm-side learning in response to the private certification.[35] Private certification may convey the same information content to firms as public certification but the information may be less credible to firms, explaining the smaller effects than public certification. 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. While we cannot conclude that there is no firm-side learning, we can reject that firm-side learning is the same from the two treatments: the public certificates almost always have statistically significantly larger effects on labor market outcomes than the private treatment.

We thus argue that the public-private-control comparison is still useful. The three-way comparison still shows, firstly, that workseekers face at least some information frictions and, secondly, that giving workseekers better technology to signal their types to firms improves their labor market outcomes more than when they are less able to credibly signal their types.

[35] An experiment 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 6: Public and Private Certification Effects 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 ≀	Index	Earnings ≀	Hourly wage ≀	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												
	Certificate Use				Search Effort					Search Effectiveness		
	Any use †	Applications ≀†	Interviews †	Offers †	Any search	Applications †≀	Hours ≀	Cost ≀	Index	Responses ‡	Offers ‡	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

Note: Coefficients are from regressing each outcome on a vector of treatment assignments, randomization block fixed effects, and prespecified baseline covariates (assessment results, self-reported assessment results, 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 ‡ (30-day recall period) or † (since treatment). Outcomes marked with ≀ use the inverse hyperbolic sine transformation. q-values control the false discovery rate within each family of outcomes.

4.3 Firms Face Information Frictions

We conduct an audit-style experiment to identify the effects of providing more information about candidates' types 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 public certifications. We vary the information available to firms when they review applications, directly replicating one stage in the job search process. This approach has the advantage of isolating the firm's response to information, without the mediation of supply-side behavior (e.g. workseekers deciding which jobs to apply to and when to use certificates).

The labor market results in the public certification experiment might occur because workseekers think firms face information frictions and hence change their search behavior. The audit study allows us to directly test if firms actually face information frictions. However, the audit experiment does not provide a comprehensive test of all possible firm responses. First, the audit study captures only initial responses to job applications, not job offers. Second, 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, online applications, which is not necessarily an important hiring channel for candidates in our sample.[36] We therefore interpret the audit experiment as testing for the existence of information frictions facing firms hiring through one potential channel, but recognize that firms may face frictions when hiring through other channels too.

We describe the experiment briefly here, with more details in Appendix G. 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 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.

We generate between six and ten applications from each candidate and randomly assign half of these to be sent with public certificates. Vacancies are randomized to receive either one or three applications with certificates identical to those used in the public certification experiment. We randomly match applications to vacancies but never send multiple applications from the same candidate

[36] Only 6% of employed candidates in our workseeker sample obtain jobs through online applications and treatment does not change this share.

to a vacancy. 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. The experimental design naturally leads to estimating the application-level equation:

$$Y_{iv} = \text{Certificate}_{iv} \cdot \beta_1 + \text{Certificate}_{iv} \cdot \text{HighIntensity}_v \cdot \beta_2 + E_{iv} + V_v + \epsilon_{iv}, \tag{2}$$

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.[37] Heteroskedasticity-robust standard errors ϵ_{iv} are clustered by vacancy, the highest level of treatment randomization, following Abadie et al. (2017). The treatment effect β_1 measures the result of using a certificate when no other applicant does so, and β_2 measures the differential effect of using a certificate when two other applicants also do so.[38] While the vacancy-level variation provides a partial test for crowd-out effects at the application level, it also allows us to examine firm-side frictions by directly estimating the vacancy-level equation:

$$Y_v = \text{HighIntensity}_v \cdot \alpha + R_v + \epsilon_v, \tag{3}$$

where HighIntensity_v is defined as before and R_v are fixed effects for the implementation round. This time, α measures the effect of assigning a vacancy to a high-intensity information treatment. This directly tests how firms respond to a richer information environment.

The analysis sample includes 3,752 applications from 632 candidates sent to 938 vacancies. We estimate treatment effects considering the full sample of vacancies (All), as well as dropping vacancies for which no application receives a response (Active). Estimates for the sample of active vacancies are reported to address concerns that vacancies may have been filled in the interim.[39]

At the application level, including a certificate makes the application 1.7 percentage points more likely to receive a response and 1 percentage point more likely to receive an interview request (Table 7, Panel A). Both effect sizes are roughly 12% of the outcome mean and significantly different from zero at the 10 percent level. This treatment effect is partly driven by the 79% of vacancies that respond to no applications.[40] When we restrict the sample to active vacancies that send a response to at least one application, the treatment effects rise to 9 and 5 percentage points, though the

[37]The vacancy-level treatment assignment is omitted because it is colinear with the vacancy fixed effects.

[38]Results are similar when we cluster standard errors by candidate, condition on candidate fixed effects, or condition on candidate-level covariates.

[39]See Neumark (2018) for a discussion about the trade-offs of using all vacancies or only vacancies that respond to at least one application in audit studies. See Bertrand and Mullainathan (2004), Rich (2014), and (Riach and Rich, 2002) for examples of the different approaches.

[40]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.

Table 7: Treatment Effects of Additional Information in Audit Study

	(1)	(2)	(3)	(4)
	Any response		Interview request	
	All	Active	All	Active
Panel A: Application-level estimates				
Certificate	0.017*	0.090*	0.010*	0.050
	(0.010)	(0.052)	(0.006)	(0.032)
Certificate × HighIntensity	-0.027*	-0.135*	-0.015	-0.072
	(0.014)	(0.069)	(0.010)	(0.045)
Outcome mean	0.137	0.706	0.087	0.455
# applications	3752	802	3752	802
Panel B: Vacancy-level estimates				
HighIntensity	0.025	0.006	0.002	-0.028
	(0.020)	(0.040)	(0.016)	(0.061)
Outcome mean	0.131	0.718	0.087	0.468
# vacancies	938	212	938	212

Note: Coefficients are from regressing each outcome on a vector of treatment assignments. Estimates at the vacancy level collapse data across applications within a same vacancy and include study-round fixed effects. Estimates at the application level include vacancy fixed effects and email address fixed effects. Heteroskedasticity-robust standard errors shown in parentheses, clustering by vacancy. The interview request outcome in columns 3-4 is a subset of the any response outcome in columns 1-2. The active sample in columns 2 and 4 exclude any vacancies where no applications receive responses.

p-value for the effect on interview responses rises to 0.119.[41]

At the vacancy level, receiving three compared to one treated application increases the firm's response rate by 2.5 percentage points and the rate of interview responses by 0.2 percentage points (Table 7, Panel B). Though none of the effects are statistically significant, the increase in response rate is meaningful–about 19% of the mean response rate. These results are consistent with firm-side frictions at this margin, but the large standard errors call for caution. When we restrict the sample to active vacancies, the treatment effects become smaller and less precisely estimated. For firms that have a non-zero response rate, receiving three applications with a certificate increases the rate of response only by 0.6 percentage points, and may even decrease the rate of interview responses by 2.8 percentage points, consistent with displacement effects observed at the application level.

[41] This result follows mechanically from the underlying econometrics. The full-sample estimates of β_1 in equation (2) are weighted averages of zero and the estimate in the sample of vacancies with any outcome variation. The weights equal the shares of vacancies respectively with and without outcome variation. A similar argument applies to β_2.

4.4 Firms Value Access to Information about Candidates' Types

If better information about candidates' types changes firms' decisions, then they should value this information. We test this claim using two incentivized choice experiments with a sample of 69 firms. We recruit the 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 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 types and obtain their contact information. See Figures A.2 and A.3 for selected screenshots of the database. 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 secondary school, complete secondary school and assessment results (all middle terciles, five middle terciles and one top tercile).

Most firms prefer candidates with higher assessment results to candidates with post-secondary education. 91% of firms rank candidates just completed secondary school and a top tercile assessment result ahead of candidates with post-secondary diplomas and all middle tercile assessment results. This shows firms substantially value the information that assessment certificates contain about workseeker types. As we discuss in Section 3.4, there is substantial heterogeneity in which specific assessment results, and hence which workseeker types, firms value most.

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 'Blinded' Certificates Have Limited Effects on Labor Market Outcomes

We interpret the certification as providing information about candidates' types to the demand and supply sides of the labor market. However, they may work through a mechanism entirely unrelated to type information. For example, certification may simply demonstrate that candidates were assessed, which firms may interpret as a positive signal about their dedication. Alternatively, certification may 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 test our preferred type information interpretation against these other two interpretations, though the test cannot distinguish between the dedication and salience interpretation we describe above.

We give some candidates blinded 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 type 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 blinded certification treatment has limited effects on labor market outcomes (Table A.12). 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 cannot reject equality of the public and blinded certification effects on most outcomes. But the small magnitude of these treatment effects favors the type 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' 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 productivity signals such as young people with limited work experience and without educational qualifications that convey accurate information about their productivity. 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. Assessing workseekers and certifying their results to both workseekers and firms 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 both workseekers and firms face information frictions. Revealing information to workseekers changes their beliefs and some measures of job search but has modest effects on their labor market outcomes. Revealing information to firms has positive effects on the response rate to applications and the interview invitation rate, though the latter effect is smaller and less precisely estimated. The distinction between workseeker- and firm-side frictions is important, as it informs how government or private firms might design information-provision products.

Our findings raise several questions for future work. We show that better information about workseekers' skills can change firms' hiring decisions. Future work could examine if incorporating this information improves firm-worker matching algorithms or algorithmic hiring rules (Abebe et al., 2018; Beam, 2016; Groh et al., 2015; Hoffman et al., 2018; Horton, 2017). We show that information on one specific set of skills is valuable. Future work could explore the many other skills that could be assessed, whose value might differ substantially across sectors and occupations.

We show that small-scale use of certifications to reduce information frictions increases employment for certified workseekers. Future work could explore the general equilibrium effects of large-scale reductions in information frictions. Small-scale certification may raise employment for certified workseekers by displacing uncertified workseekers outside our sample, in which case the employment effect of large-scale certification may be smaller or even zero.[42] Our experiment is not designed to evaluate this possibility. But our results offer some suggestions. The audit study shows that certification effects are attenuated, although not to zero, when more applications are accompanied by better information. This suggests decreasing returns are possible as more workseekers are certified. The workseeker study shows that better information can increase match quality, proxied by earnings conditional on employment. This provides a mechanism for lower information frictions to increase aggregate employment, by raising the value of some matches to firms above minimum or reservation wages. Even if alleviating information frictions has no effect on aggregate employment, it may still raise workseeker or firm welfare by reducing job search and vacancy posting costs and reducing the frequency of bad hires that lead to separation (Donovan et al., 2018).

We show that improved information is valuable to firms and potentially workseekers. Future work could explore the prospects for market-based provision of certification services. Our findings

[42] Some studies of large-scale active labor market programs find smaller effects at larger scale (Lise et al., 2004; Crépon et al., 2013). But Blundell et al. (2004) find no displacement effects of a large-scale job search assistance and wage subsidy program.

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 US$160 search subsidy on certification (compared to 24% on training and 27% on transport). This may inspire questions about why agents do not already use certification services in the market. Some large firms in South Africa and other countries already use psychometric assessments in hiring (Autor and Scarborough, 2008). For small firms that make infrequent hires, developing in-house assessment systems is unlikely to be cost-effective. Entry by third-party firms is difficult because the fixed costs of developing assessment tools and establishing credibility may be high. In ongoing work with Harambee, we are exploring the financial viability of third-party assessment services.

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.

Ahn, S., R. Dizon-Ross, and B. Feigenberg (2019): "Improving Job Matching Among Youth," Working Paper, Columbia University.

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

Anderson, M. (2008): "Multiple Inference and Gender Differences in the Effects 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.

ATTANASIO, O., A. KUGLER, AND C. MEGHIR (2011): "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 3, 188–220.

AUTOR, D. AND D. SCARBOROUGH (2008): "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments," *Quarterly Journal of Economics*, 123, 219–277.

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 (2009): "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.

BLUNDELL, R., M. C. DIAS, C. MEGHIR, AND J. VAN REENEN (2004): "Evaluating the employment impact of a mandatory job search program," *Journal of the European Economic Association*, 2, 569–606.

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 Affect Economic Preferences, Strategic Behavior, and Job Attachment," *Proceedings of the National Academy of Sciences*, 106, 7745–7750.

Burks, S., B. Cowgill, M. Hoffman, and M. Housman (2015): "The Value of Hiring Through Employee Referrals," *Quarterly Journal of Economics*, 130, 805–839.

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.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013): "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment," *Quarterly Journal of Economics*, 128, 531–580.

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 Effect: 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-Efficacy, Executive Function, and Temporal Discounting in Kenya," *Behaviour Research and Therapy*, 101, 30–45.

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

Garlick, R., K. Orkin, and S. Quinn (2019): "Call Me Maybe: Experimental Evidence on Frequency and Medium Effects 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.

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 Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment," *Journal of Labor Economics*, 35, 345–285.

ILO (2016): "Enabling Environment for Sustainable Enterprises in South Africa," International Labour Office – Geneva.

Ioannides, Y. M. and L. D. Loury (2004): "Job Information Networks, Neighborhood Effects, 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.

Jepsen, C., P. Mueser, and K. Troske (2016): "Labor Market Returns to the GED Using Regression Discontinuity Analysis," *Journal of Political Economy*, 124, 621–649.

Kahn, L. (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.

Lam, D., C. Ardington, and M. Leibbrandt (2011): "Schooling as a Lottery: Racial Differences 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.

Lise, J., S. Seitz, and J. Smith (2004): "Equilibrium policy experiments and the evaluation of social programs," Working paper 10283, National Bureau of Economic Research.

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.

McKenzie, D. (2017): "How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence," *World Bank Research Observer*, 32, 127–154.

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

Pallais, A. (2014): "Inefficient Hiring in Entry-Level Labor Markets," *American Economic Review*, 104, 3565–3599.

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

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 Deficits, 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 Certificate (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 Efficiency 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 Economic Forum (2018): "The Future of Jobs Report," World Economic Forum – Geneva.

A Conceptual Framework

We consider a labor market facing two-sided information frictions: firms and/or workseekers may imperfectly observe workseekers' types. 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, type T_i, and homogeneous time endowment, E. We begin with a scalar type so workseekers are vertically differentiated, but the same intuition applies with multidimensional types that allow workseekers to be horizontally differentiated. 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 $U^W(T_i, W_i)$, discussed in more detail below, while time spent on other activities yields utility $U^O(T_i, O_i)$. These utility functions include pecuniary and non-pecuniary benefits and costs such as wages, monetary costs of job search, and leisure.[43] She then receives a job offer with probability $p(T_i, W_i^*)$ with an associated wage $Y(T_i, W_i^*)$. 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 $O_i^* = E - W_i^*$ to equate the marginal return from job search and the marginal return from other activities

$$\frac{\partial U^W\left(T_i, W_i^*(T_i)\right)}{\partial W_i} = \frac{\partial U^O\left(T_i, E - W_i^*(T_i)\right)}{\partial W_i}. \tag{4}$$

If she receives a job offer, she accepts any offer with wage $Y(T_i, W_i^*) > \bar{Y}(T_i, W_i^*)$. Her decision decisions may depend on her heterogeneous type T_i in two ways. First, $W_i^*(\cdot)$ will be a non-degenerate function of T_i if type 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 types, in which case the optimal time allocation to searching for formal work will be increasing in type. Second, her wage and/or reservation wage will depend on skill if firms differentially value different types (discussed below) or workseekers with different types 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

[43] 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

costs of capital. Effective labor $L^E = \sum_{i \in f} T_i$ is the sum of the types of all workers employed by the firm. Absent information frictions, firms' optimal choice of labor satisfies $p \times \frac{\partial Q_f\left(L^{E*}, K^*\right)}{\partial L} = w$, where w is the wage. Workers are paid their marginal revenue product, so workers with higher types receive higher wages. This model gives rise to a threshold rule, where each firm hires workers in a type 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 $Y^*(T_i, W_i * (T_i))$ in which all workseekers who receive job offers accept them.

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

$$\mathbb{E}\left[\frac{\partial U^W\left(T_i, W_i^*(S_i)\right)}{\partial W_i}\right] = \mathbb{E}\left[\frac{\partial U^O\left(T_i, E - W_i^*(T_i)\right)}{\partial W_i}\right] \tag{5}$$

where the expectation is taken over the distribution of ϵ_i. The optimal time allocations in conditions (4) and (5) will be equal if both utility functions are linear in type 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' types. The concavity of the CES production means that the expected marginal revenue product from each work seeker is lower than their expected type:

$$\mathbb{E}\left[p \times \frac{\partial Q_f\left(L^{E*}, K^*\right)}{\partial L^{E*}}\right] < p \times \frac{\partial Q_f\left(\mathbb{E}\left[L^{E*}\right], K^*\right)}{\partial L^{E*}}. \tag{6}$$

If wages are completely flexible and all firms face information frictions, then wages will fall for each type. 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.[44] 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

[44] 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.

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 types, respectively $\tilde{T}_i = T_i + \epsilon_i$ and $\hat{T}_i = T_i + \nu_i$. 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 types and because they know firms cannot assess their types, increasing the risk of 'incorrect' job and wage offers. Firms' offer decisions will be distorted because they do not know workseekers' types and because they know workseekers cannot assess their own types, 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.

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 certification experiment. However, 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 workseekers and only firms.

The framework does not generate a simple prediction about which workseekers will be affected most by information frictions. The incidence of distortions depends on the joint distribution of true types and perceived types and on the parameterization of the model. For example, if firms observe types with classical measurement error and shrink observed types toward the population mean, then information frictions will attenuate the return to higher type. Alleviating frictions will raise high-type workseekers' wages and potentially employment and will hurt low-type workseekers' wages and potentially employment. Alternatively, the wedge between true and observed types may be larger at low levels of true skills, perhaps because high-type 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 types. But the core results are unchanged if types are multidimensional and at least one dimension is imperfectly observed. This increases the scope for variation in match quality.

With multidimensional types, information frictions can lower wages for all workers.

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 types, 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 gradually observe workers' productivity 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' productivity (Arcidiaono et al., 2010; Lange, 2007).

B Further Details on Assessments

B.1 Assessments

We assess workseekers' performance in six domains. Detailed information on all six assessments is available at `https://www.assessmentreport.info`, including sample questions. The numeracy, communication 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. This assessment is used by one of South Africa's largest retail chains as part of their in-house screening process for new recruits.

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.

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. This assessment uses 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. The 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. The 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 scores have 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).

The first 23% of the workseekers were assessed using self-reported measures of *control* and *flexibility* instead of the focus and planning assessments. These assessments used 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.

These assessments are widely used in research and in some hiring processes. However, we conducted several validation exercises to verify that they performed well in this population before beginning the certification process.

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).

All assessments were administered by registered industrial psychologists employed or contracted by Harambee. Psychologists approved the design of the certificates, oversaw each assessment session, and delivered briefings to candidates to interpret results. This ensures compliance with South African law on psychometric testing in workplace settings. See https://www.hpcsa.co.za/Uploads/editor/UserFiles/ for details.

The assessment process is designed to avoid excess stress or exhaustion for candidates. Assessments are split over two days to avoid exhaustion. 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.2 Firms' Use of Assessments

The numeracy, communication and concept formation assessments have used by Harambee for several years to select candidates for further job readiness training. Harambee has placed over 20,000 candidates in entry-level jobs using these tests since 2011. We show descriptive data on use of these assessments by 33 large firms in retail, hospitality, logistics and corporate services in Table A.1. Firms can select which assessments they use. All firms used at least one cognitive test to screen candidates: 24 used all three tests, 2 used two and 7 used one. In contrast, only 20 (60%) required a CV, 19 (57%) required a matric certificate with test scores and 1 required references. This suggests firms find this skill information as, if not more informative, than other traditional signals of skill for selecting candidates. Harambee also administers a set of career aptitude measures provided by a psychometric testing firm. 22 (67%) of firms in this sample used this assessment score to screen applicants, suggesting they value horizontal differentiation. We did not include this assessments in

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

Note: This table captures use of psychometric and other assessment scores by 33 Harambee client firms. The assessments are described in Appendix B. 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.

the certification because it is a proprietary instrument whose psychometric properties we could not assess.

C 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 adapt a method from Attanasio et al. (2011) to decompose 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:

$$\begin{aligned}
&\underbrace{\mathbb{E}[Earn|Treat=1]-\mathbb{E}[Earn|Treat=0]}_{\text{ATE for earnings}} \qquad (7)\\
=\;& \underbrace{\left(\mathbb{E}[Earn|Treat=1,Work=1]-\mathbb{E}[Earn|Treat=0,Work=1]\right)}_{\text{ATE for earnings}\,|\,\text{employment}}\cdot \underbrace{Pr[Work=1|Treat=1]}_{\text{Treated employment rate}}\\
+\;& \underbrace{\mathbb{E}[Earn|Treat=0,Work=1]}_{\text{Control earnings}\,|\,\text{employment}}\cdot\underbrace{\left(Pr[Work=1|Treat=1]-Pr[Work=1|Treat=0]\right)}_{\text{ATE for employment}}.
\end{aligned}$$

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.[45]

All terms in equation (7) 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 (7). 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 with low employment rates, where latent earnings are unobserved for most candidates.

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

D Heterogeneous Outcome Levels and Treatment Effects by Workseeker Type

Analyzing the relationship between workseeker types and other candidate attributes, including treatment effects, is difficult because we conduct six assessments, generating a multidimensional measure of type. In this appendix we explain the challenge of estimating heterogeneous treatment effects by type 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 type.

There are two obvious parametric approaches to testing for heterogeneous treatment effects by type, each of which has serious limitations. First, we could collapse multidimensional assessment results into a single index and then regress outcomes on treatment assignments, the single index, and their interaction. However, this imposes the assumption that any treatment effect heterogeneity over assessment results is linear and additive, ruling out the possibility of complementarity across domains. Second, we could regress outcomes on treatment assignments, each assessment result, and all possible interactions. However, this requires estimating a large number of parameters and may

[45] Attanasio et al. (2011) show that the intensive margin effect can be further decomposed into two terms: the treatment effect on earnings conditional on candidates' baseline characteristics, and the difference in baseline characteristics between employed candidates in the treatment and control groups. However, neither of these terms is point identified.

Table A.2: Heterogeneity in Labor Market Treatment Effects by Skill Using Single Indices

	(1)	(2)	(3)	(4)
Outcome	Employed	Employed	Employed	Employed
Index	Share top terciles	Share top - share bottom terciles	PC_1(Scores)	$\hat{\text{Pr}}$(Employ|Scores)
Treatment	0.052***	0.052***	0.052***	0.052***
	(0.012)	(0.011)	(0.011)	(0.012)
Treatment × index	0.032	0.013	0.004	-0.039
	(0.057)	(0.028)	(0.025)	(0.026)
Mean outcome	0.309	0.309	0.309	0.309
# observations	6607	6607	6607	6607
# clusters	84	84	84	84

Coefficients are from regressing each outcome on a vector of treatment assignments, the relevant index, treatment × index interactions, randomization block fixed effects, and prespecified baseline covariates (assessment results, self-reported assessment results, 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.

require imposing some rule to aggregate over potentially different signs.

The first approach generates no evidence of heterogeneity using four different indices A.2. The first two indices are scores based on the number of assessments in which candidates achieve top and bottom terciles. The third index is the first principal component of the cardinal assessment results. The fourth index is a weighted average of the cardinal assessment results, with weights derived from a control group regression of endline employment on the assessment results and interactions between them. The third index weights the different assessments based on their covariance; the fourth assessment weights the different assessments based on their relationship with employment. The result is consistent across all four approaches: no large or statistically significant heterogeneity by assessment results.

Given the limited heterogeneity using single indices, we also propose 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 assessments. Let the average treatment effect for this subset of candidates be Δ_M. If treatment effects are increasing in assessment results, then Δ_M should be larger than the average treatment effect for candidates scoring in the lowest tercile for all assessments, Δ_L. Similarly, Δ_M should be smaller than the average treatment effect for candidates scoring in the highest tercile for all assessments, Δ_H. More generally, the average treatment effect for candidates with any assessment result should be higher (respectively lower) than the average treatment effect for candidates with strictly lower (respectively higher) results in all domains.

More formally, define

$$\Delta_{a_1,\ldots,a_J} = \mathbb{E}\left[Y_i(1) - Y_i(0) | A_{i,1} = a_1, \ldots, A_{i,J} = a_J\right] \tag{8}$$

as the average treatment effect for candidates scoring in terciles $a_1, \ldots, a_J$ for the J assessments. This expectation is taken over the distibution of candidate-level heterogeneity conditional on assessment results. Define

$$\Delta^+_{a_1,\ldots,a_J} = \Delta_{a_1,\ldots,a_J} - \mathbb{E}\left[\Delta_{\tilde{a}_1,\ldots,\tilde{a}_J}\right] \tag{9}$$

equal the difference in the average treatment effect between candidates with assessment results $a_1, \ldots, a_J$ and candidates with assessment results $\tilde{a}_1, \ldots, \tilde{a}\tilde{a}_J$, 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 possible combinations of terciles that are strictly lower than s_j in all domains j. Define

$$\Delta^-_{a_1,\ldots,a_J} = \Delta_{a_1,\ldots,a_J} - \mathbb{E}\left[\Delta_{\tilde{a}_1,\ldots,\tilde{a}_J}\right] \tag{10}$$

analogously, where $\tilde{a}_j > a_j, \forall j$.

If treatment effects are increasing in assessment results, then $\Delta^+_{a_1,\ldots,a_J} > 0 > \Delta^-_{a_1,\ldots,a_J}$. 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

$$\begin{aligned} H_+ : \Delta^+_{a_1,\ldots,s_J} &= 0 \\ H_- : \Delta^-_{a_1,\ldots,s_J} &= 0 \\ H_d : \Delta^+_{a_1,\ldots,s_J} - \Delta^-_{a_1,\ldots,a_J} &= 0. \end{aligned}$$

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

We implement this test by estimating the average treatment effect within each assessment result 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 assessment result 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 there are only a few assessments.

We do not find strong evidence of heterogeneous certification effects on labor market outcomes by assessment results (Table A.3). Certification effects on employment and hours are slightly higher for candidates with lower assessment results. Certification effects on earnings, wages, and contract status do not vary by assessment results.

These results show that there is no robust evidence of heterogeneous treatment effects by assessment results. We suggest two interpretations for this result. First, the assessments primarily differentiate candidates horizontally, rather than vertically. Very few candidates obtain all top or all bottom terciles and most candidates obtain some top and some bottom terciles (Section 2.5). The single indices capture relatively little of the joint variation in assessment results: 37% for the

Table A.3: Heterogeneity in Labor Market Treatment Effects by Assessment Results Using Dominance Test

	(1) Employed	(2) Hours ≀	(3) Earnings ≀	(4) Hourly ≀ wage	(5) Written contract
Treatment effect relative to units with strictly					
Lower assessment results	-0.005	-0.074	0.156	0.133	0.010
	(0.029)	(0.100)	(0.149)	(0.083)	(0.012)
Higher assessment results	0.051***	0.178**	0.105	0.049	0.000
	(0.024)	(0.090)	(0.148)	(0.094)	(0.013)
p: Differences equal	0.139	0.065	0.801	0.495	0.571

Note: This table reports estimates of treatment effect heterogeneity using dominance tests. To implement these tests, we estimate the average treatment effect within each cell defined by the Cartesian product of the terciles of all assessment results. We estimate the difference between the cell-specific treatment effect and the treatment effects in all cells with strictly *lower* terciles on all assessments (rows 1-2). We estimate the difference between the cell-specific treatment effect and the treatment effects in all cells with strictly *higher* terciles on all assessments (row 3-4). If the former quantity is larger than the latter, then treatment effects are increasing by assessment results. 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 ≀ use the inverse hyperbolic sine transformation.

principal component and less for the other indices. The dominance test compares groups of candidates with strictly dominated or strictly dominating assessment results, but only 31% of candidates score higher than another candidate in all assessments and only 30% of candidates score lower than another candidate in all assessments.

Second, there is substantial variation in demand for different types of workseekers. Our incentivized choice experiment shows that firms' relative ranking of performance in different domains is very heterogeneous. This means that the type of variation firms value is unlikely to be detected by tests based on single indices or dominance approaches.

We can use this dominance approach to test if the levels of baseline variables differ by assessment results. We simply define

$$\bar{Y}_{a_1,\ldots,a_J} = \mathbb{E}\left[Y_i | A_{i,1} = a_1, \ldots, A_{i,J} = a_J\right] \tag{11}$$

instead of $\Delta_{a_1,\ldots,a_J}$ and then define $\bar{Y}^-_{a_1,\ldots,a_J}$ and $\bar{Y}^+_{a_1,\ldots,a_J}$ analogously.

E Additional Results Discussed in Paper

E.1 Summary Statistics and Balance Tests

This section reports summary statistics for the baseline workseeker sample (Table A.4), endline workseeker sample (Table A.5), and sample of workseekers used in the audit study (Table A.6). Balance tests for equal means of baseline measures are also reported in the final column of Table A.4.

Table A.4: 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 qualification	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
Offers 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 overconfident	6873	0.503	0.352			0.588
Fraction of assessments underconfident	6873	0.115	0.208			0.367

Note: Table shows summary statistics for selected baseline variables. Percentiles are omitted for binary variables. All monetary figures are reported in South Africa Rands. 1 Rand $\approx$ 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.5: 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
Offers 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 overconfident	6605	0.345	0.237		
Fraction of assessments underconfident	6605	0.176	0.166		

Note: Table shows summary statistics for selected endline variables. Percentiles are omitted for binary variables. All monetary figures are reported in South Africa Rands. 1 Rand $\approx$ 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.6: 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 certificate	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

Table A.7: 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 certification treatment effects 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 effect	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

Note: 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.

E.2 Benchmarks Magnitudes of Earnings Effects

In this section we show that the earnings effects are substantial relative to two local benchmarks.

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 firms and unions agreed minimum wages for the sector which were often then applied to all firms in the sector (Budlender et al., 2015; Isaacs, 2016). Table A.7 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.

The average treatment effect on earnings is equal to roughly one sixth of the adult poverty line or 6-10% of the monthly minimum wage at the time of the study.

E.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 differ by treatment arm (Table A.8). Non-response does not differ 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.8: 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

Note: Coefficients 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.9: Non-response by Baseline Covariates Group in Each Post-Treatment Survey Round

	(1)	(2)
	Text Message Survey	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 coefficients jointly zero	0.000	0.018
Mean outcome	0.170	0.040
# observations	6889	6889
# clusters	84	84

Note: Coefficients are from regressions of round-specific 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.

Table A.10: Treatment Effects on Labor Market Outcomes Without Covariates

	(1)	(2)	(3)	(4)	(5)
Panel A: Employment Status Measures					
	Employed	Month 1	Month 2	Hours ≀	Index
Treatment	0.046***	0.038***	0.057***	0.175***	0.130***
	(0.013)	(0.012)	(0.014)	(0.058)	(0.029)
q: Treatment effect = 0	0.001	0.002	0.001	0.002	
Mean outcome	0.309	0.465	0.437	8.848	0.001
Mean outcome for employed				28.847	
# observations	6607	6604	6607	6598	6609
# clusters	84	84	84	84	84
Employment 'Quality' Measures					
	Earnings ≀	Hourly wage ≀	Written contract	Index	
Treatment	0.336***	0.206***	0.018*	0.107***	
	(0.076)	(0.041)	(0.010)	(0.029)	
q: Treatment effect = 0	0.001	0.001	0.029		
Mean outcome	159.291	9.840	0.120	0.006	
Mean outcome for employed	518.291	32.283	0.392		
# observations	6589	6574	6575	6609	
# clusters	84	84	84	84	

Note: Coefficients are from regressing each outcome on a vector of treatment assignments and randomization block fixed effects. 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 ‡ (30-day recall period) or † (since treatment). Outcomes marked with ≀ use the inverse hyperbolic sine transformation.

E.4 Additional Treatment Effects

Table A.11: Treatment Effects on Self-Beliefs through Time

	(1)	(2)	(3)	(4)
	Numeracy correct		Above-median self-esteem	
Public	0.233***	0.315***	0.001	-0.001
	(0.013)	(0.015)	(0.013)	(0.015)
Private	0.200***	0.333***	-0.002	0.016
	(0.015)	(0.016)	(0.014)	(0.015)
p: public = private	0.010	0.240	0.806	0.239
Mean outcome	0.396	0.399	0.553	0.479
# observations	6601	5297	6609	5027
# clusters	84	84	84	84
Data source	Endline	Text messages	Endline	Text messages

Coefficients are from regressing each outcome on a vector of treatment assignments, randomization block fixed effects, and prespecified baseline covariates (assessment resultss, self-reported assessment results, 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. Above-median self-esteem is an indicator equal to one if the candidate's response on a shortened version of the Rosenberg (1965) self-esteem scale is above the sample median. Numeracy correct is an indicator if the candidate's self-reported tercile rank in numeracy equals their actual rank. Columns (1) and (3) report results from the main phone follow-up survey. Columns (2) and (4) report results from the text message survey conducted 2-3 days after treatment. We do not adjust these results for multiple testing because we measure two conceptually different outcomes.

Table A.12: Treatment Effects on Labor Market Outcomes

	(1) Employed	(2) Month 1	(3) Month 2	(4) Hours ≀	(5) Index	(6) Earnings ≀	(7) Hourly wage ≀	(8) Written contract	(9) 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

Note: Coefficients are from regressing each outcome on a vector of treatment assignments, randomization block fixed effects, and prespecified 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 ‡ (30-day recall period) or † (since treatment). Outcomes marked with ≀ use the inverse hyperbolic sine transformation. q-values control the false discovery rate within each family of outcomes.

F Certificates Used in Skill-blind Treatment

Figure A.1: Sample Skill-Blind Certificate

REPORT ON CANDIDATE COMPETENCIES
-Personal Copy-

This report contains results from the assessments you took at Harambee in Phase 1 and Phase 2. These results can help you learn about some of your strengths and weaknesses and inform your job search.

You completed assessments on English Communication (listening, reading and comprehension) and Numeracy today in Phase 2. In Phase 1, you completed a Concept Formation assessment which asked you to identify patterns.

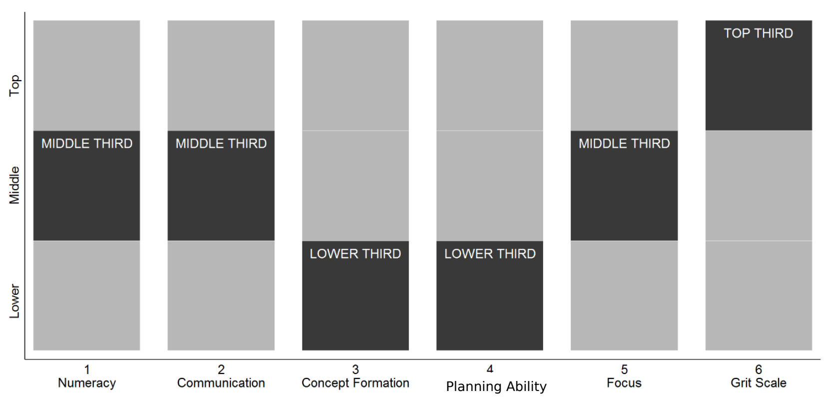
1. The Numeracy tests measure various maths abilities. Your score is the average of the two maths tests you did today at Harambee.
2. The Communication test measures English language ability through listening, reading and comprehension.
3. The Concept Formation test measures the ability to understand and solve problems. Candidates with high scores can generally solve complex problems, while lower scores show an ability to solve less complex problems.

You also did some games and questionnaires to measure your soft skills:

4. The Planning Ability Test measures how you plan your actions in multi-step problems. Candidates with high scores generally plan one or more steps ahead in solving complex problems.
5. The Focus Test looks at your ability to pick out which information is important in confusing environments. Candidates with high scores are able to focus on tasks in distracting situations.
6. The Grit Scale measures candidates' determination when working on difficult problems. Candidates with high scores spend more time working on the problems rather than choosing to pursue different problems.

Your results have been compared to a large group of young South African job seekers who have a matric certificate, are from socially disadvantaged backgrounds and have been assessed by Harambee.

You scored in the MIDDLE THIRD of candidates assessed by Harambee for Numeracy, MIDDLE THIRD for Communication, LOWER THIRD for Concept Formation, LOWER THIRD for Planning Ability, MIDDLE THIRD for Focus and TOP THIRD for the Grit Scale.



DISCLAIMER
Please note that this is a confidential assessment report and is intended for use by the person specified above. Assessment results are not infallible and may not be entirely accurate.

Notes: This figure shows an example of the certificates given to candidates in the skill-blind treatment group. The certificates 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 certificates and guidelines on how to request more certificates.

G 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 1-8 are already completed (Appendix Table A.13) and round 9 is 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.13: Implementation Details of Audit Study Rounds 1 to 8

	Rounds 1 to 8	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8
Panel A: Search intensity									
Candidates invited	2,220	204	378	270	234	234	270	270	360
Candidates responded (all)	632	66	126	68	76	71	87	52	86
Panel B: Audit study									
Vacancies	1,018	148	195	101	110	131	118	106	109
Applications	3,992	591	777	404	437	524	472	387	400
Responses received	555	55	130	37	89	56	76	55	57

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 25 percent of the work seekers contacted across all rounds responded to our message within a week. Work seekers in the final sample of the audit study are slightly selected. As observed in Appendix Table A.14, compared to workseekers in the supply-side study they are more likely to be male, and to have post-school qualifications or less than high school. They perform slightly worse in the literacy tests, and better in concept formation and grit. Importantly, they are more likely to have worked in the past 7 days, and to have been assigned to the private treatment arm in the supply-side study.[46]

[46] However, note that the audit study uses within-applicant randomization, so the average treatment of providing more information to firms for the sample of audit study participants is identified without adjusting for selection.

Table A.14: 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	0.90	0.30	632	-	-	-
Includes a cover letter	0.13	0.30	632	-	-	-
Includes a copy of ID document	0.50	0.50	632	-	-	-
Includes information about high-school completion	0.59	0.49	632	-	-	-
Panel B: Characteristics of workseekers						
Public treatment	0.31	0.46	632	0.33	0.47	6,891
Private treatment	0.37	0.48	632	0.31	0.46	6,891
Age	23.3	3.15	632	23.7	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-highschool certificate	0.24	0.43	632	0.21	0.41	6,891
Completed highschool	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/communications 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
Grit assessment score (z score)	0.11	1.00	632	0.03	0.99	6,891
Worked in the last 7 days (endline)	0.41	0.49	632	0.38	0.48	6,891

We process the applications received and record information on when the application was received, 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 effect 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 benefits, or discriminate based on appearance, race, or gender. The curated list rarely exceeds 200 vacancies per round. Appendix Table A.15 Panel A, describes the sample of vacancies. Typical sectors include sales, admin, call center, industrial, restaurant, and service.

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 different 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.[47] We generate between 6 and 10 applications per work seeker in each round completed to date, so that the total number

[47] 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 filled during that wait period.

Table A.15: Vacancy-Level Attributes

	Mean	Std Dev.	# Obs
Panel A: Job sector			
Sales	0.48	0.50	1018
Admin	0.21	0.41	1018
Call centre	0.11	0.32	1018
Industrial	0.09	0.29	1018
Restaurant	0.04	0.20	1018
Service	0.03	0.17	1018
Uncategorized	0.17	0.38	1018
Panel B: Responses to applications submitted			
Response to any application	0.14	0.35	1018
Response to all applications	0.07	0.26	1018
Response missing	0.08	0.27	1018

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.

For the analysis, we drop the entire block of applications submitted to the 8 percent of vacancies for which at least one application is missing a response (Appendix Table A.15 Panel B). Only vacancies with complete application response data are included in the analysis. This does not induce a sample selection problem or a loss of treatment orthogonality, since blocks of applications are randomly assigned to a vacancy. The resulting sample includes 3,752 applications from 632 candidates sent to 938 vacancies.

Table A.16: Descriptive Statistics for Application-Level Attributes

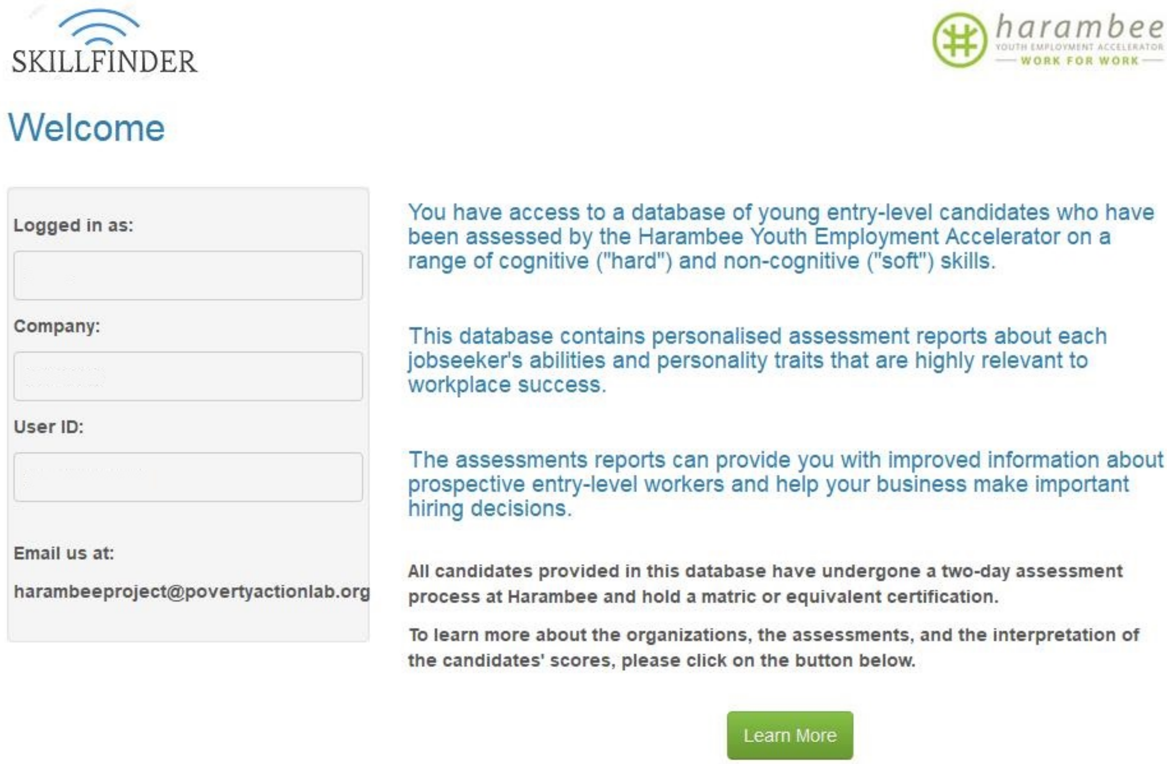
	Mean	Std Dev.	# Obs
Panel A: Characteristics of applications submitted			
Had one report in a vacancy with one report	0.12	0.33	3,752
Had one report in a vacancy with three reports	0.38	0.48	3,752
Had no report in a vacancy with one report	0.37	0.48	3,752
Had no report in a vacancy with three reports	0.13	0.33	3,752
Panel B: Responses to applications submitted			
Any response received	0.14	0.35	3,752
Interview request received	0.09	0.28	3,752
Acknowledgement received	0.02	0.12	3,752
More information requested	0.03	0.18	3,752
Scam, rejected, closed	0.002	0.046	3,752
Panel C: Responses conditional on any response received			
Interview request	0.61	0.49	534
Acknowledgement	0.10	0.30	534
More information	0.23	0.42	534
Scam, rejected, closed	0.02	0.12	534

As shown in Appendix Table A.16, roughly half of a workseeker's applications in the analysis sample 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, 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 (9 percent of the full sample of applications).

The audit study design described above allows us to directly identify the average effect of alleviating pure demand-side frictions. However, it only identifies such effect 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 offers 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.

H Screenshots of Platform Used in Firm Willingness-to-Pay Measurement

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



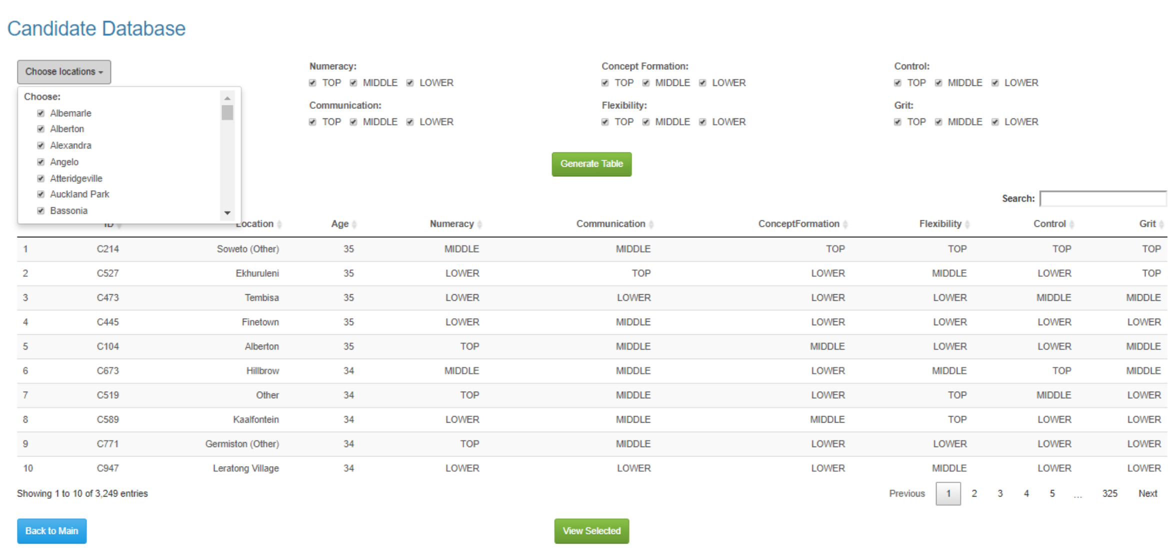


Figure A.3: Screenshot of Individual Candidate Profile on Platform

THE WORLD BANK

Candidate Information

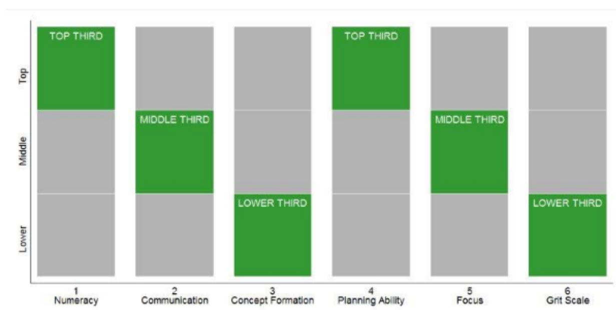
ID Number: CID67929

Age: 34

Location: Other

Date of Assessments: 2016-12-06

The candidate obtained the following assessment results:



The candidate completed assessments in Numeracy, English Communication (listening, reading, comprehension), and Concept Formation:

1. The Numeracy test measures a candidate's ability to apply numerical concepts at a National Qualifications Framework (NQF) level, such as working with fractions, ratios, money, percentages and units and performing calculations with time and area. This score is an average of two numeracy tests the candidate completed.
2. The English Communication test measures a candidate's grasp of the English language through listening, reading and comprehension. It assesses at an NQF level, for example measuring the ability to recognise and recall literal and non-literal text.
3. The Concept Formation Test is a non-verbal measure that evaluates a candidate's ability to understand and solve problems. Those with high scores are generally able to solve complex problems, while lower scores indicate an ability to solve less complex problems.

The candidate also completed standardised questionnaires to assess their soft skills:

4. The Flexibility Scale measures whether candidates actively consider several approaches to solving a problem. Those with high scores generally explore several avenues to find the best possible solution, while low scores indicate considering fewer approaches. **In lieu of the Flexibility scale, some candidates will have the Planning scale listed. Flexibility and Planning should be used interchangeably.**

5. The Control Scale measures whether candidates react impulsively or systematically when faced with problems. Candidates with high scores generally deal with problems systematically, while those with lower scores tend to react spontaneously. **In lieu of the Control scale, some candidates will have the Focus scale listed. Control and Focus should be used interchangeably.**

6. The Grit Scale measures whether candidates show determination when working on challenging problems. Those with high scores generally spend more time working on challenging problems, while those with low scores choose to pursue different problems.

Ready to contact the candidate? Click to view contact information:

View Contact Information | Print this page