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

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

February 21, 2020

Abstract

Workseekers' job search decisions and firms' hiring decisions use potentially noisy information about workseekers' skills. We show that assessing workseekers' skills in multiple domains, giving workseekers the assessment results, and helping them to credibly share the results with firms increases workseekers' employment and earnings. It also alters their beliefs and search behavior. Giving information only to workseekers has similar effects on beliefs and search, but substantially smaller effects on employment and earnings. Giving information only to firms increases callbacks and interview invitations. These patterns are consistent with both firms and workseekers facing information frictions that distort search and hiring.

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

*We are grateful for helpful comments from Peter Arcidiacono, Paolo Belli, Emily Breza, Rulof Burger, Stefano Caria, Jishnu Das, Giacomo di Giorgi, Taryn Dinkelman, Erica Field, Simon Franklin, Marek Hanusch, Rachel Heath, Clement Imbert, Lisa Kahn, Melanie Morten, Simon Quinn, Imran Rasul, Duncan Thomas, Gabriel Ulyssea, Chris Woodruff, and seminar participants at the NBER Summer Institute, Barcelona Summer Forum, Duke, Georgia, Notre Dame, NYU, Oxford, UIUC, and the World Bank. We appreciate exceptional research assistance from Wim Louw, Wendy Trott, Shelby Carvalho, Allegra Cockburn, Goodwill Mpofu, Emmanuel Bakirdjian and Nilmini Herath in South Africa and Lukas Hensel, Svetlana Pimkina, Laurel Wheeler, Gabriel Cunha and Rachel Sayers at our institutions. Thanks to staff at the Harambee Youth Employment Accelerator; the Abdul Latif Jameel Latif Poverty Action Lab Africa Office; and our institutions. The project is conducted in collaboration with the World Bank Jobs Group and Africa Region Gender Innovation Lab, 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 UCT Commerce Faculty Ethics in Research Committee, the Duke Institutional Review Board for the Protection of Human Subjects in Non-Medical Research (protocol # D0368) and the Oxford Department of Economics Research Ethics Committee (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

Many decisions in the labor market rely on potentially noisy information about workseekers' skills. Firms decide if and whom to hire, at what wage, typically using information on education and past work experience. Workseekers decide if and how they will search based on feedback on their performance in education and in past work. This information may only weakly predict performance in specific jobs. Limited information for firms can lead to hiring poorly-matched workers and to wedges between wage offers and productivity (Altonji and Pierret, 2001; Arcidiaono et al., 2010; Farber and Gibbons, 1996; Kahn and Lange, 2014). Limited information for firms can also reduce average wage offers and reduce employment under some production technologies or if firms face uninsured risks from hiring mistakes (Aigner and Cain, 1977; Pallais, 2014). Workseekers with limited information about their skills may search for jobs that are a poor match for their skills or withdraw from search entirely (Conlon et al., 2018; Mueller et al., 2018; Spinnewijn, 2015). These search distortions may lead to lower wages and employment.

We study whether providing new information about workseekers' skills affects job search and hiring decisions and workseekers' outcomes in the labour market. We assess 6,891 workseekers' skills in six domains using standardised assessments. The assessments draw on existing tools used by job placement agencies and large firms and cover general workplace skills such as communication, grit, and numeracy. The workseekers are drawn from a population where information frictions may be important. They are unemployed or underemployed youths in urban South Africa with limited post-secondary education, work experience, and access to referral networks. We separately manipulate firms' and workseekers' information about workseekers' skills. This allows us to test separately for firm- and workseeker-side information frictions, and their implications for job search and hiring.[1]

We first show that giving workseekers their assessment results in a form they can easily and credibly share with firms improves the workseekers' labor market outcomes. To show this, we randomly select some workseekers for a 'public' certification intervention. This intervention gives them electronic and physical certificates showing their assessment results. The certificates show the workseekers' names and national identity numbers, are branded by the widely known agency that conducts the assessments and the World Bank, describe the assessments, and show the workseekers' assessment results. We track job search and employment for both publicly certified workseekers and a control group of workseekers who receive no certificates and do not learn their assessment results. In the three to four months following certification, publicly certified workseekers use certificates in job applications, update beliefs about their skills, and target their job search toward jobs that they think value their skills. Their employment rate increases by 17% (5 percentage points), weekly earnings by 34%, and hourly wages by 20%. The rise in earnings reflects both higher employment

[1] We use the term 'information friction' to refer to limited information about workseekers' skills throughout the paper. Our experiments are not designed to speak to other types of limited information that may be present in the labor market market, such as limited information about workers' effort and or about vacancy characteristics.

and higher earnings conditional on employment.

This first experiment shows that information frictions exist but does not show who has limited information about workseekers' skills: firms, workseekers, or both. We show that workseekers face information frictions by randomly selecting some workseekers for a 'private' certification intervention. This intervention gives them a single physical certificate that shows their assessment results and a description of the assessments, but excludes all identifying information and branding that might make the certificate a credible source of information to firms. Private certification has the same effects as public certification on workseekers' beliefs and search targeting, a positive effect on earnings that is smaller than the public certification effect, and no effect on employment. These results show that giving workseekers more information about their skills changes their behavior and outcomes in the labor market, but less than when they can also share the information with firms. This suggests that both workseekers and firms have limited information about workseekers' skills.

We run an audit-style experiment as a further check for firm-side information frictions. This manipulates firms' information without any scope for changes in workseeker behavior. We submit applications to real job vacancies using real resumes from workseekers in our sample. We submit multiple applications per vacancy, randomizing whether applications include public certificates. Applications including certificates get 13% more callbacks and 11% more interview invitations. This response is consistent with firms having limited information about workseekers' skills and acquiring more information from skill certification. We do not observe hiring decisions in the audit study, so we view this as secondary evidence relative to our experiments with workseekers.

These three experiments are designed to answer our primary question: to test if firms, workseekers, or both sides of the market have incomplete information about workseekers' skills. In addition, we present secondary analyses to further characterise how information frictions affect this labour market. These use heterogeneity analysis and smaller experiments and hence should be interpreted more cautiously than our primary results. First, learning specific assessment results is important, not just learning that workseekers have been assessed. This shows that the certification effects are not driven by firms using workseekers' decisions to get assessed as a signal for tenacity or proactivity or by firms basing hiring decisions purely on the certificates' branding. Second, in this labor market, preferences for different skills vary across firms and relative performance in different assessments varies across workseekers. This pattern is more consistent with horizontal than vertical differentiation: certification helps firms identify which workseekers are suited for specific jobs, rather than identify a subset of workseekers suited for all jobs (consistent with Lise and Postel-Vinay, 2020). Third, certification has larger effects on the labor market outcomes of workseekers who lack other ways to communicate their skills to employers, like work experience and university education. Fourth, the positive effect of certificates on callbacks in the audit study does not persist when we send multiple applications with certificates to the same vacancy. This result is consistent with the effects of more information about workseekers' skills attenuating at scale. However, our other results are consistent with economic mechanisms that predict positive effects at scale and attenuated effects on callbacks do not necessarily lead to attenuated effects on hiring (Jarosch and Pilossoph, 2019).

Our primary contribution is to show that both firms and workseekers face information frictions and that the effects of information provision are driven by responses on both sides of the labor market. Knowing whether frictions affect either or both sides of the labour market is important for understanding how private actors or government might address information frictions. Interventions targeted at one side of the market are common, but may not be optimal in the presence of two-sided frictions. For example, some job search assistance programs or careers counsellors offer skill assessments to workseekers. This can inform workseekers and improve their search targeting. But if the assessment results are not certified to firms and firms face information frictions, then their hiring choices and wage offers will remain distorted and workseekers' improved search targeting will have limited returns. On the other hand, some firms use skill assessments as part of their hiring process (Autor and Scarborough, 2008; Hoffman et al., 2018). This can inform firms and improve their hiring decisions. But if workseekers face information frictions, then firm-side assessments will not help they apply for jobs that match their skills. Firms can only assess workseekers who apply for their vacancies, so they may be assessing a pool that omits the best-match applicants. Some labor market platforms like LinkedIn have begun to offer assessment systems that provide information to both sides of the market. These may be preferable to one-sided information mechanisms if both firms and workseekers face information frictions. But the question of who faces information frictions is still important for designing these two-sided information mechanisms, such as deciding which side(s) of the market to charge for assessment and certification.

We contribute to a broad body of microeconomic research on information frictions in the labor market. Researchers have shown that workseekers' skills are not perfectly observed by firms or workseekers and have studied mechanisms that provide information about workseekers' skills: formal education qualifications, referrals from network connections, performance evaluations from past employers, and skill assessments.[2] However, previous studies have either examined frictions on only one side of the market or have examined simultaneous information revelation to both sides of the market. For example, one literature studies how firms learn about their own workers' skills by observing their performance over time, abstracting away from workseeker-side information frictions (Altonji and Pierret, 2001; Arcidiaono et al., 2010; Farber and Gibbons, 1996; Kahn and Lange, 2014). Research on referrals studies how firms learn about workseekers' skills, abstracting away workseeker-side learning about their skills. Altmann et al. (2018), Ahn et al. (2019), and Belot

[2] Alfonsi et al. (2017), Clark and Martorell (2014), Jepsen et al. (2016), and MacLeod et al. (2017) study the information content of workseekers' formal education qualifications. Ioannides and Loury (2004) review the literature on network referrals and Beaman and Magruder (2012), Beaman et al. (2018), Pallais and Sands (2016), and Heath (2018) examine the information referrals convey about workseekers' skills. Hardy and McCasland (2017) show that firms may use internships as a screening device to assess prospective workers' skills.

et al. (2018) show that workseekers have limited information about their labor market prospects but do not examine firm-side frictions. Our work is closest to recent papers that experimentally manipulate information about workseekers' skills. Abebe et al. (2019) and Bassi and Nansamba (2019) study what happens when both workseekers and firms can see skill assessment results, while Abel et al. (2019) and Pallais (2014) study what happens when both workseekers and firms can see performance evaluations from workseekers' past employers. These papers show that information revelation changes workseekers' outcomes and interpret this as evidence of firm-side information frictions. But none of these papers separately manipulate firms' and workseekers' information, so their experiments are not designed to separately identify firm- and workseeker-side frictions.[3]

The distinction between firm- and workseeker-side information frictions is also important for modeling macroeconomic implications of information frictions. Models with information frictions can explain macroeconomic patterns of labour market dynamics (Donovan et al., 2018). However, both empirical and theoretical work in this area generally focuses on only one side of the market: either firms' limited information about workseekers' skills (Doppelt, 2016; Kennan, 2010) or workseekers' limited information about the job arrival rate or wage offer distribution they face (Belot et al., 2018; Conlon et al., 2018; Falk et al., 2006; Gonzalez and Shi, 2010). Our findings can help to inform the development of models with two-sided information frictions. Our finding that workseekers have imperfect information about their measured skills can help to provide an explanation for why workseekers in some of these models have uncertain and heterogeneous beliefs about the job arrival rate or wage offer distribution they face.

Our work is also relevant to the extensive literature on active labor market programs, reviewed by Card et al. (2018), Heckman et al. (1999), and Kluve et al. (2019), amongst others. We show a skill assessment and certification intervention can substantially improve participants' employment and earnings.[4] The employment effect is almost three times larger than the mean standardized effect size of the active labor market programs reviewed by Card et al. (2018). Skill assessment and certification may be a useful addition to alternative mechanisms for learning about workseekers' skills. It is available to first-time workseekers, unlike reference letters or performance evaluations from past employers. Assessment results can be certified to multiple employers, while workplace performance at one employer may be imperfectly observed by other employers (Kahn, 2013). Certification may be cheaper than formal educational qualifications. It may help workseekers excluded from referral networks or firms who receive referrals based on factors poorly aligned with workseekers' skills

[3]Abel et al. (2019) run two experiments, one that shows reference letters to only firms and one that shows reference letters to both firms and workseekers. But they do not manipulate workseekers' information conditional on firms' information and they do not observe a common outcome across across the two experiments, limiting their ability to speak to workseeker-side frictions.

[4]In a similar spirit, several papers show that ALMPs' effectiveness can be increased when they make low-cost changes that provide more information to firms and/or workseekers (Abel et al., 2019; Belot et al., 2018; Wheeler et al., 2019). This relates to a broader literature studying the effects of changes to the design of active labor market programs (McCall et al., 2016).

(Beaman and Magruder, 2012; Beaman et al., 2018).

We describe the economic environment in Section 2: a simple conceptual framework, the context we study, our sample, and the skill assessments. In Section 3, we describe the public skill certification experiment and the treatment effects on workseekers' labor market outcomes. In Section 4, we analyze the roles of firm- and workseeker-side frictions. In Section 5 we discuss the secondary results about what workseekers and firms learn from skill certification, what this implies for the effects of certification on different types of workseekers, and what this might imply for general equilibrium effects of a richer information environment. We conclude in Section 6 and briefly discuss questions around markets for assessment-based certification.

2 Economic Environment

2.1 Conceptual Framework

In this section, we sketch a static framework to illustrate how either workseeker- or firm-side information frictions can lower two labor market outcomes: the employment rate and that mean wage conditional on employment. We use the framework to illustrate the mechanisms linking frictions to labor market outcomes and guide our empirical work.

Consider a stylized economy consisting of infinitely many type W_1 and W_2 workseekers and type J_1 and J_2 jobs. Workseekers choose to not search, search for type 1 jobs, or search for type 2 jobs. Searching for either type of job incurs fixed cost $C > 0$. A type i workseeker searching for type j jobs meets a firm offering such as job with probability $P_{i,j}$. Conditional on meeting, the workseeker earns wage $W_{i,j}$ wage and produces output $V_{i,j} \geq W_{i,j}$. The workseeker receives utility $U\left(P_{i,i} \cdot W_{i,i}\right) - C$ if she searches and zero otherwise, implying that she has a reservation wage $\underline{W}_i$ that depends on the search cost.[5] Non-employment is possible if search costs are high relative to the expected utility of working (which leads to some non-participation) or if the meeting probability $P_{i,j}$ is less than one for some (i, j). To simplify the discussion that follows, we assume fraction p of all workseekers and all jobs are type 1 and that workseekers are horizontally differentiated: type i workseekers are better at searching for type i jobs, produce the most output in type i jobs, and earn the highest wages in type i jobs. Under these assumptions, type i workseekers always prefer searching for type i jobs rather than type j jobs. But none of the ideas in the framework depend on these assumptions.

Limited information about workseekers' skills can enter this environment in two ways. First, we consider the case where workseekers observe their types with error and firms observe workseekers' true types. This type of friction can occur if workseekers receive noisy information about their own type from education or work experience or if they have little education or work experience. With this type of friction, each workseeker chooses whether and where to search based on her

[5] For simplicity, we assume that firms post and commit to wages before workseekers make search decisions. This implies that all workseekers who choose to search for type j jobs will accept them if offered.

perceived type. If a type i workseeker searches for the 'wrong' type of job, she is less likely to meet a firm and, conditional on meeting a firm, will earn a lower wage and produce less. This type of friction reduces mean wages conditional on employment by generating some workseeker-job type mismatches. This can also reduce the employment rate through two mechanisms: workseekers who search for the wrong type of jobs are less likely to meet firms and mismatched workseeker-job types may not generate enough output to meet the wage floor. The former mechanism is likely if different types of firms hire using different channels, like posting formal adverts versus hiring walk-ins. The latter mechanism is likely if search costs, and hence reservation wages, are high or if there is a legal minimum wage. Belot et al. (2018) and Falk et al. (2006) prove results of this flavor formally.

Second, workseekers may observe their true types, while firms observe workseekers' types with error. This type of friction can occur if attributes observable to firms, like educational qualifications or past work experience, are noisy proxies for skills. With this type of friction, workseekers search for the 'right' types of jobs but firms will not know the type of the workseekers they meet. If type j firms believe that fraction q of the workseekers they meet are type j, then the expected output from each hire is $q \cdot V_{j,j} + (1-q) \cdot V_{i,j}$. If firms' utility is a concave function of their output, then they will offer a wage lower than $q \cdot W_{j,j} + (1-q) \cdot W_{i,j}$. Concavity can arise from firms' production technology or from uninsured risks from bad hires. Possible uninsured risks include lost customers or damaged equipment from hiring the 'wrong' workseekers and then incurring severance pay and dispute resolution costs when firing these workseekers. This reduces mean wages conditional on employment and, if the offered wage is below the reservation wage or minimum wage, reduces the employment rate. Aigner and Cain (1977) and Jovanovic (1979) prove results of this flavor formally.

This simple framework shows that either firm- or workseeker-side information frictions can lower the employment rate and mean wage conditional on employment. Hence, observing that employment and/or wages rise when both firms and workseekers acquire more information does not show whether firms or workseekers face information frictions. This highlights the importance of the separate variation we generate in firms' and workseekers' information sets. 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 an interaction effect. We focus on the static case for simplicity, but recognize that the effect of information frictions may differ in a dynamic framework with learning by firms or workseekers (Farber and Gibbons, 1996; Conlon et al., 2018; Lange, 2007).

The framework allows either horizontal or vertical differentiation. We define horizontal differentiation as type i workseekers being more productive than type j workseekers in type i jobs and vice versa. We define vertical differentiation as type i workseekers being more productive than type j workseekers in all jobs. In both cases, either firm- or workseeker-side information frictions lower the employment rate and the mean wage conditional on employment. With horizontal differentiation, frictions on either side of the market may lower wages conditional on employment for all workseekers. With vertical differentiation, firm-side frictions can increase wages for type j workseekers if they are mistaken for type i workseekers.

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. First, information frictions are likely. Grades and grade progression in most primary and secondary schools are weakly correlated with independently measured skills (Lam et al., 2011; Taylor et al., 2011; Van der Berg and Shepherd, 2015). This limits the information 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 examination grades weakly predict performance in post-secondary education and firms report in interviews that the grades convey limited information about skills (Schöer et al., 2010).[6] 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.

Second, 'bad' hires are costly. Firing a worker requires a complex and lengthy process and can be challenged by even temporary employees in courts and specialized dispute resolution bodies. Probationary work is permitted but regulated and probation periods cannot exceed three months (Bhorat and Cheadle, 2009). Firms report challenges understanding labor regulation, contributing to the perceived cost of separations.[7] Giving firms free consulting on labor regulation increases hiring, consistent with perceptions of regulation constraining employment (Bertrand and Crépon, 2019).

Third, reservation and minimum wages exist. Minimum wage compliance in the formal sector is high (Bhorat et al., 2016; ILO, 2016). Commute costs are high and likely to raise reservation wages (Kerr, 2017). The nearly universal state pension system gives workseekers in multigeneration households access to non-labor market income, which also increases reservation wages (Abel, 2019).

Fourth, employment rates are low. 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] Low employment in the

[6] The limited information content of education qualifications is consistent with the large role of referrals in hiring, with more than half of all firms and two thirds of small firms reporting this as their preferred recruitment mechanism (Schöer et al., 2014).

[7] Firms with less than 50 employees report an average of two dispute resolution cases in the previous year, requiring an average of 11 days of staff time per case (Rankin et al., 2012). 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).

[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. Unemployment rates exclude those

presence of information frictions, costs from bad hires, and reservation wages is consistent with our conceptual framework. Particularly low employment for youths is also consistent with information frictions, as youths have less search and work experience to reveal their types. Many other factors can contribute to low employment rates; we merely argue that a role for information frictions is plausible.

2.3 Sample Recruitment and Data Collection

We recruit a sample of 6,891 young, actively searching people from low-income backgrounds with limited work experience. This is not a population-representative sample. But it does represent a theory-relevant population likely to face information frictions. Workseekers in our sample have limited access to traditional ways to learn about their skills and communicate their skills to prospective employers: university education, work experience, or access to referral networks. We recruit only active workseekers, so we do not examine the relationship between information frictions and labor market participation decisions.

To recruit the sample, we work with the Harambee Youth Employment Accelerator, a social enterprise that "builds solutions to address a mismatch of demand and supply in the South African youth labor market by connecting employers with inexperienced workseekers." Harambee recruits candidates through radio and social media advertising and door-to-door recruitment in low-income neighborhoods. Interested candidates register online and complete a phone-based screening questionnaire.[9] Eligible candidates are invited to two days of standardized skill assessments. A small share of candidates are invited to further job readiness training based on their assessment results and residential location, but we show in Section 3 that this is irrelevant to our study. Our sample consists of all candidates who arrive at Harambee for the second of these two testing days, on 84 operational days.[10]

We conduct three surveys to measure workseekers' labor market outcomes, search, and beliefs about their skills and the labor market. The baseline is a self-administered but supervised questionnaire on desktop computers at Harambee. This is administered after candidates have done skills assessments but before they receive any information about their results. We collect endline data in a 25-minute phone survey roughly 3-4 months after treatment.[11] The phone survey response rate is 96%, leaving an endline sample of 6,607 respondents. The response rate is balanced across treatment groups (Table C.6) and unrelated to most baseline covariates (Table C.7). We also conduct

in full-time education or not in the labor force.

[9] Candidates are eligible to work with Harambee if they are aged 18-29, have legal permission to work in South Africa, have completed secondary school, have at most twelve months of formal work experience, have no criminal record, and are from disadvantaged backgrounds. This information is self-reported but checked against administrative data for some candidates.

[10] 1.4% are invited to further job readiness training based on their assessment results and residential location, but we show in Section 3 that this is irrelevant to our study.

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

a short text message survey 2-3 days after treatment. Respondents receive mobile phone airtime payments for answering the text message and phone surveys.

2.4 Job Search and Employment in Our Sample

This section describes relevant patterns around labor market outcomes and job search in our sample. We report summary statistics for key baseline and endline variables for the 6,891 workseekers in Tables C.1 and C.2. Respondents are 99% Black African, 62% female, and on average 24 years old. 17% have a university degree or diploma, 21% have some other post-secondary certificate, and 99% have completed secondary school. Relative to the population of the study province aged 18-29, our sample is more educated, more female, and more likely to be actively searching for work (Table C.4).

38% of the sample worked in the week before the baseline, 70% had ever worked before. Conditional on working, mean weekly earnings in the week before the baseline was 90 USD PPP (565 South African Rands), slightly below the minimum wage for a full-time worker in most sectors. At endline, wage work was eight times more common than self-employment. Most work was relatively short-term, with median and mean tenures of 2 and 7 months respectively.

97% of the sample searched for work in the week before the baseline. In that week they spent on average 39 USD PPP (242 South African Rands) and 17 hours searching. They submitted on average 10 applications in the preceding month and received 1.2 offers, though the medians for both measures are zero. The job search and application process is somewhat formal: 38% of the candidates employed at endline reported that they submitted written applications for their current job and 47% reported that they had a formal interview, though 48% also reported using a referral.

2.5 Assessments

We conduct six assessments with workseekers: communication, concept formation (similar to a Raven's test), focus, grit, numeracy, and planning. Firms have demonstrated interest in the results of these assessments, though they obviously also use other information in hiring decisions. The numeracy assessment was developed by a large retailer. Harambee has screen roughly 160,000 prospective workers for paying client firms using these assessments. Appendix A describes each assessment in detail, their psychometric properties, and other research documenting associations with workplace productivity.

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. Table C.1 shows standardized scores on all six assessments and Table C.3 shows the correlation matrix of the skills. There is a fairly even spread of candidates over the distribution and little evidence of ceiling effects.

We interpret candidates with different assessment results as different worker types, in the language of the conceptual framework. Scores are weakly correlated across assessments, with pairwise correlations between 0.05 and 0.51. Hence, the assessments mainly horizontally differentiate candidates based on their relative skills rather than ranking them in a single dimension.

Candidates have inaccurate beliefs about their own types. We ask candidates in which tercile they believe they ranked for each of the communication, concept formation, and numeracy assessments after taking the assessments but before any candidates learn their results. Only 8% of candidates answer correctly for all three assessments and 28% of candidates answer incorrectly for all three assessments. Overconfidence is more common than underconfidence: 22% of candidates overestimate their tercile on all three assessments and 1% underestimate their tercile all three assessments.

3 Labor Market Effects of Certification

3.1 Intervention

Our first certification intervention gives candidates information about their assessment results and allows them to share the results with prospective employers. Candidates receive a certificate describing the assessments and their performance (Figure 1). They receive 20 color copies printed on high-quality paper and an email version. Each certificate briefly describes Harambee and its placement and assessment work, describes the assessments, and shows the tercile in which the candidate ranked on each assessment. The certificate describes the background of the assessed candidates (high school graduates aged 18-34 from disadvantaged backgrounds) to help readers interpret the terciles.[12] The certificate directs the reader to https://www.assessmentreport.info/ for more information on Harambee and the assessments. The website shows sample questions for each assessment and describes how psychologists have designed and evaluated the assessments. To link candidates with certificates, each certificate shows the candidate's name and unique national identity number. National identity numbers are typically shown on resumes and school transcripts in South Africa. To provide credibility to the assessments and results, the certificate is branded with the World Bank logo and Harambee logo. The latter brand is widely recognized in South African marketing surveys (Mackay, 2014).

Each candidate receives their certificates during a group briefing with a psychologist. The psychologist explains what each assessment measures and how to interpret the results on the certificate. They explain that workseekers can but do not have to attach the certificate to future job applications and that they can request more certificates from Harambee. The research team and Harambee psychologists jointly developed a briefing script and PowerPoint presentation. Research assistants monitored each briefing to ensure psychologists used the script.

[12] We piloted versions of the certificates containing only rankings, only cardinal scores, and both rankings and cardinal scores. Both workseekers and firms preferred certificates with only rankings, as they could not easily interpret cardinal scores.

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.

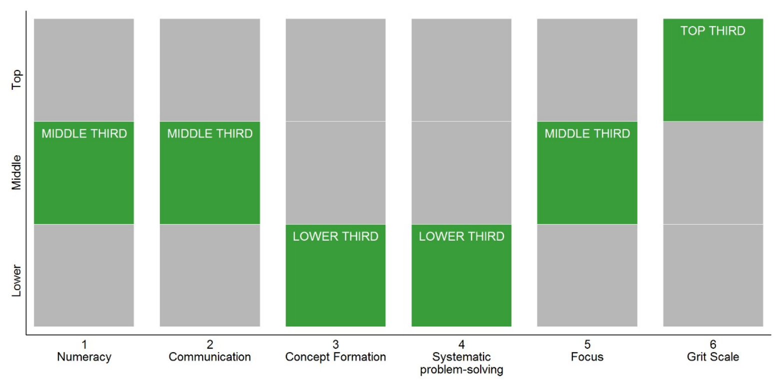
«**name**» completed assessments on English Communication (listening, reading, comprehension), Numeracy, and Concept Formation:

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 workseeker received 20 printed certificates, an email certificate, and guidelines on how to request more certificates.

In terms of the conceptual framework, this certification intervention gives information directly to workseekers. Workseekers then choose whether to share this information with firms. The effects of the intervention may reflect reductions in firm- or workseeker-side information frictions. In either case, the framework predicts that certified workseekers will have higher employment and higher earnings conditional on employment.

3.2 Experimental Design

We randomly divide our workseeker sample into a certification group, a control group, and other groups discussed in the next section. We randomize treatment by assessment date to reduce risks of spillovers between treated and control workseekers, assigning 2,247 workseekers over 27 days to certification and 2,274 workseekers over 27 days to control. Randomization is sequential and stratified, with days randomized within blocks of 6-10 upcoming days. Table C.1 shows that the randomization generates balanced treatment assignments. Treated workseekers receive the certification intervention described above. Control workseekers receive no information about their assessment results and no assistance sharing results with firms. All treated and control workseekers receive roughly one hour of job search counselling before the assessments on how to create an email address and how to prepare and dress for an interview. They also receive an email with a CV template, interview tips, and job search tips.[13] This differs from the design in Abebe et al. (2019), where treated workseekers receive skill certification and job search counselling while control workseekers receive neither.

We estimate treatment effects using models of the form

$$Y_{id} = \mathbf{T}_d \cdot \Delta + \mathbf{X}_{id} \cdot \Gamma + S_d + \epsilon_{id}, \tag{1}$$

where Y_{id} is the outcome for workseeker i assessed on date d, $\mathbf{T}_d$ is a vector of treatment assignments, $\mathbf{X}_{id}$ 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. All labor market and job search measures use 7-day recall periods, except where we specify otherwise. 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. We thus analyze treatment effects on realized outcomes, rather than latent outcomes that may be non-zero for the non-employed or non-searching. We also estimate quantile treatment effects on selected outcomes, which allows us

[13]Harambee invites some workseekers for further training and job search assistance. These invitations depend partly on their assessment results and may only be issued months after assessment. By the endline, only 1.4% of our sample are invited for further interaction with Harambee and only 0.17% receive a job offer through their further interaction with Harambee. These outcomes are uncorrelated with treatment status and all our results are robust to dropping these workseekers.

Table 1: Treatment Effects on Labor Market Outcomes

	(1) Employed	(2) Hours[c]	(3) Earnings[c]	(4) Hourly wage[c]	(5) Written contract
Treatment	0.052	0.201	0.338	0.197	0.020
	(0.012)	(0.052)	(0.074)	(0.040)	(0.010)
Mean outcome	0.309	8.848	159.291	9.840	0.120
Mean outcome for employed		28.847	518.291	32.283	0.392
# observations	6607	6598	6589	6574	6575
# clusters	84	84	84	84	84

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. Mean outcome is for the control group. All outcomes use a 7-day recall period unless marked with [a] (30-day recall period) or [b] (since treatment). Outcomes marked with [c] use the inverse hyperbolic sine transformation. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

to focus on the distribution of outcomes for only employed or only searching candidates.

The estimating equations and variable definitions are prespecified. We report treatment effects on some outcomes that are not prespecified but note when we do so. In Appendix C we show that our results are robust to adjusting for multiple testing and omitting the prespecified covariates $\mathbf{X}_{id}$.

3.3 Certification Improves Labor Market Outcomes

The first main effect of certification is to increase employment. Current employment rises by 5.2 percentage points from a control group mean of 30.1 percentage points (Table 1 column 1). We also ask about employment in each calendar month after treatment and show in Table C.11 that certification increases employment in every month after treatment. The effect on current employment is substantial: almost three times larger than the mean standardized effect size of active labor market programs reviewed by Card et al. (2018), larger than the effect of another South African intervention that helped workseekers get reference letters from past employers (Abel et al., 2019), and similar to the effect of a program that subsidized firms to hire workseekers from similar backgrounds (Levinsohn et al., 2013).

Certification increases average weekly hours by 20% (column 2). We code hours worked as zero for non-employed candidates. So the treatment effect on hours may reflect two effects: an extensive margin effect as treatment increases the employment rate and an intensive margin effect as treatment increases the hours that employed candidates work. We adapt a decomposition proposed by Attanasio et al. (2011) to identify these two effects (details in Appendix B). We define the extensive margin effect as the treatment effect on employment multiplied by mean hours worked for employed control group candidates. Intuitively, this is the rise in hours we would see if treatment increased employment but the marginally and inframarginally employed treated candidates worked the same

Table 2: Treatment Effects on Labor Market Outcomes at Extensive and Intensive Margins

	(1) Hours[c]	(2) Earnings[c]	(3) Hourly wage[c]	(4) Written contract
Total effect	0.201	0.338	0.197	0.020
	(0.052)	(0.073)	(0.039)	(0.010)
Extensive margin	0.189	0.269	0.141	0.020
	(0.042)	(0.059)	(0.031)	(0.005)
Intensive margin	0.013	0.069	0.056	-0.000
	(0.020)	(0.040)	(0.028)	(0.008)
Treatment effect conditional	0.036	0.195	0.159	-0.001
on employment	(0.058)	(0.113)	(0.078)	(0.024)

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. Treatment group employment is 36%, so the conditional effects on all outcomes are roughly three times larger than the corresponding intensive margin effect. Heteroskedasticity-robust standard errors are shown in parentheses, clustering by treatment date. All outcomes use a 7-day recall period unless marked with [a] (30-day recall period) or [b] (since treatment). Outcomes marked with [c] use the inverse hyperbolic sine transformation.

average hours as the inframarginally employed untreated candidates. We define the intensive margin effect as the difference between the treatment effect on hours and the extensive margin effect on hours. The entire effect on hours is explained by the extensive margin effect (Table 2 column 1). This shows that treated candidates do no work longer hours conditional on employment, they are simply more likely to be employed.

The second main effect of certification is to increase earnings. Weekly earnings increase by 34% (Table 1 column 3). The increase in earnings is an economically meaningful change, equal to 17% of the weekly adult poverty line in South Africa (details in Appendix C.2). 27 percentage points of the 34% increase in earnings is explained by the rise in employment (the extensive margin effect, shown in Table 2 column 2). This implies an intensive margin effect on earnings of 7 percentage points per candidate. Hourly wages, calculated by dividing earnings by hours, also increase by 20% (Table 1 column 4). The extensive and intensive margins account for respectively 14 and 6 percentage points of the 20% increase (Table 2 column 3).

These results are consistent with the conceptual framework: more information about workseeker skills (i.e. types) increases both employment and mean earnings conditional on employment. The results are not consistent with a special case of the framework where more information increases job-finding rates but has no effect on the output of firm-worker matches. The results are consistent with more information increasing match output, so that more latent matches generate enough value for firms and workers to be worth making. This explanation also matches the quantile treatment effects on earnings. These are positive throughout the earnings distribution, though not always significantly different to zero (Figure 2).

Finally, certification increases by 2 percentage points the probability of having a written contract, Statistics South Africa's definition of a formal job (Table 1 column 5). This effect is entirely explained by the higher employment rate (Table 2 column 4).[14]

4 Separating Workseeker- and Firm-side Information Frictions

Certification may increase employment and earnings by providing information to firms, to workseekers, or both sides of the market. As we explain in the introduction, this distinction matters for designing government or market-based remedies to information frictions.

In this section, we show that both sides of the market face information frictions, suggesting that the effects of certification are explained by both sides of the market acquiring new information. Our argument proceeds in three steps. First, we show that public certification changes workseekers' beliefs and search behavior in ways that are consistent with either or both of firm- and workseeker-side information frictions. Second, we discuss another arm of our workseeker experiment that reveals information only to workseekers. The results of this intervention, are consistent with both firm- and workseeker-side information frictions existing and are not consistent with only one-sided frictions. Third, we discuss an audit-style experiment that reveals information only to firms. The results of this experiment are consistent with a role for firm-side information frictions.

4.1 Certification Changes Job Search and Beliefs

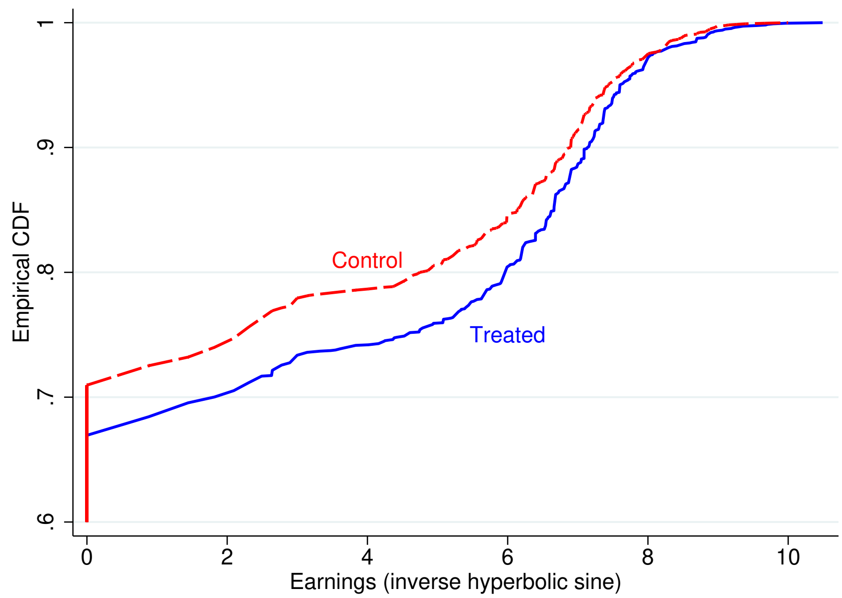
We document three patterns in the effects of certification on workseekers' beliefs and job search behavior. First, certification changes workseekers' beliefs about their skills. We ask candidates if they think they scored in the top, middle, or bottom third on each of the six assessments. Certification increases the fraction of assessments where candidates' self-assessments match their measured results from 0.39 to 0.45 (Table 3 column 1).[15] In contrast, certification has no effect on candidates' generalized self-esteem. This shows that their updated beliefs about the skills do not lead to more general updating about their self-worth (column 2).

[14] Of the 5.2 percentage point increase in employment, 4 percentage points are into wage employment and 1.2 into self employment. The wage and self employment measures are not prespecified outcomes.

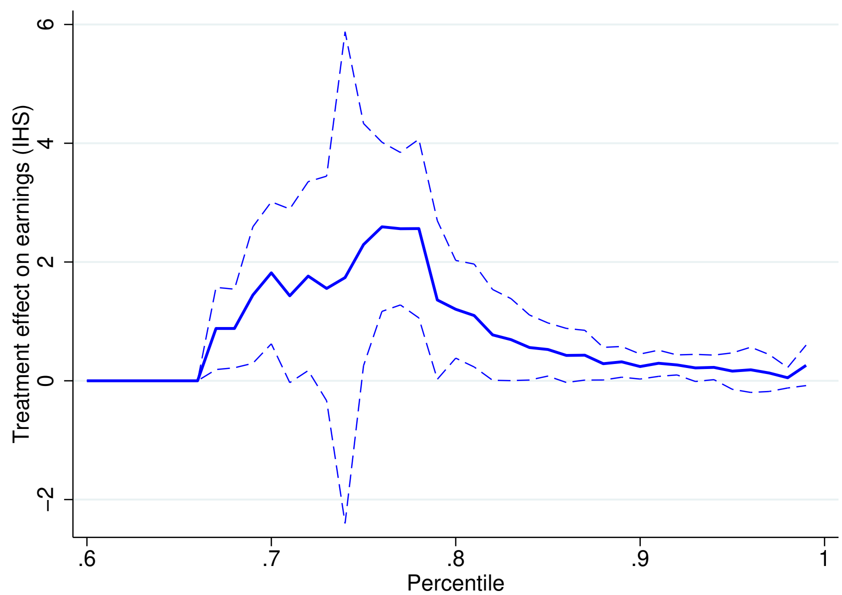
[15] This question measures candidates' beliefs about their assessment results. These may differ from their beliefs about their domain-specific skills, if for example they believe the assessments are noisy measures of their skills. To address this possibility, we ask candidates if their communication and numeracy skills are in the top, middle, or bottom third of people in a reference group (ages 18-34, from disadvantaged backgrounds, with high school education). This is not a question about their result on a specific assessment. We ask only about communication and numeracy skills because candidates are more likely to understand what these skills mean than the other four skills we assess. Using this measure of beliefs about skills, treatment increases the share of the two skills where candidates' beliefs match their assessment results by 12.4 percentage points (standard error 2.2 p.p.). This shows that candidates' updated beliefs are not assessment-specific. We collect this measure only for a random 50% sample of the first 3,000 candidates to complete the survey and then dropped it to save survey time. Treatment also reduces the fraction of assessments for which candidates are overconfident or underconfident about their performance. We do not report detailed results on other ways in which beliefs update, as this is not the primary focus of the paper.

Figure 2: Quantile Treatment Effects on Earnings

Panel A: Empirical Distributions of Earnings in Control and Public Certification Groups



Panel B: Quantile Treatment Effects of Public Certification on Earnings



Note: Panel A shows the empirical distributions of earnings in the control and public certification groups. Earnings are the inverse hyperbolic sine transformation of South African Rands, with 1 Rand ≈ 0.16 US$ in purchasing power parity terms. Earnings are coded as zero for candidates who are not working. The vertical axis in Panel A is truncated below at the 60$^{\text{th}}$ percentile because earnings below that value are zero. The vertical Panel B shows the quantile treatment effects (QTEs) of public certification. These are unconditional QTEs, estimated without controlling for any covariates or stratum fixed effects. The 95% pointwise confidence intervals allow heteroskedasticity and clustering by treatment date. The confidence intervals exclude zero at all percentiles except 73-74, 86, and 93-99.

Table 3: Public and Private Certification Effects on Beliefs, Search, and Labor Market Outcomes

	(1) Skill belief accurate	(2) > median self-esteem	(3) Targeted search		
Public certification	0.158	0.001	0.052		
	(0.008)	(0.013)	(0.010)		
Private certification	0.123	-0.002	0.047		
	(0.008)	(0.014)	(0.010)		
p: public = private	0.000	0.806	0.698		
Mean outcome	0.389	0.553	0.155		
# observations	6607	6609	6609		
# clusters	84	84	84		
	(4) Used report[b]	(5) Applications with report[b,c]	(6) Interviews with report[b]	(7) Offers with report[b]	(8) Expected offers[a,c]
Public certification	0.699	1.682	0.432	0.112	0.106
	(0.013)	(0.040)	(0.023)	(0.011)	(0.019)
Private certification	0.289	0.572	0.144	0.036	0.053
	(0.012)	(0.033)	(0.017)	(0.008)	(0.023)
p: public = private	0.000	0.000	0.000	0.000	0.025
Mean outcome	0.000	0.000	0.000	0.000	4.198
# observations	6609	6598	6597	6597	6531
# clusters	84	84	84	84	84
	(9) Worked	(10) Hours[c]	(11) Earnings[c]	(12) Hourly wage[c]	(13) Written contract
Public certification	0.052	0.201	0.338	0.197	0.020
	(0.012)	(0.052)	(0.074)	(0.040)	(0.010)
Private certification	0.011	0.066	0.162	0.095	0.017
	(0.012)	(0.048)	(0.078)	(0.046)	(0.009)
p: public = private	0.002	0.011	0.028	0.030	0.769
Mean outcome	0.309	8.848	159.291	9.840	0.120
# observations	6607	6598	6589	6574	6575
# clusters	84	84	84	84	84

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. Mean outcome is for the control group. Skill belief accurate is the share of the six assessments where the candidate's perceived tercile matches their actual tercile. 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 scored highest. 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. All outcomes use a 7-day recall/forecast period unless marked with [a] (30-day recall/forecast period) or [b] (since treatment). Outcomes marked with [c] use the inverse hyperbolic sine transformation. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

Second, changes the types of jobs that candidates target. 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 strictly highest from 0.16 to 0.21 (column 3).[16]

Third, candidates use certificates in job applications (columns 4-7). 70% of candidates use the certificates with at least one job application between treatment and endline, 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.

The first two patterns suggest a role for workseeker-side frictions: candidates update their beliefs about their skills and apply to a different mix of jobs based on the updated beliefs. The third pattern suggests a role for firm-side information frictions: candidates use reports with job applications, making the applications more informative to employers, leading to more job interviews and offers. Some combination of these patterns leads candidates to expect 11% more offers in the next month, from a control group mean of 4.2 offers (column 8).

Taken together, these three patterns are consistent with certification reducing both firm- and workseeker-side information frictions. However, these three patterns are not sufficient to show whether reducing only one of these frictions could generate the employment and earnings effects of certification. For example, the revised search targeting might by itself increase employment and earnings, as firms might ignore the certificates attached to job applications. Alternatively, the certificates attached to job applications might by themselves increase employment and earnings, as the revised search targeting might be ineffective. We therefore run two more experiments that separately manipulate the information available to firms and workseekers.[17]

Before proceeding to the next experiments, we note that certification does not change multiple measures of job search effort in the month before the endline: probability of searching, number of applications submitted, hours spent searching, and money spent on search (Table C.11). There are two possible explanations for this pattern. First, certification may change how workseekers search – targeting different jobs and using certificates in applications – without changing their search effort. This is consistent with a special case of the conceptual framework where information frictions change how firms and workseekers match but do not change the share of workseekers who choose to search. Second, certification may temporarily change search effort but the endline may occur too late to detect this change. 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 C.11). This suggests that any changes in worseeker behavior that increase employment must occur soon after treatment. The

[16]This search targeting measure is not prespecified. The result is similar for the fraction of candidates searching for jobs that most value the assessment in which they *think* they scored highest.

[17]We could alternatively explore this using back-of-the-envelope calculations or a more formal model. For example, if roughly half of job offers are accepted and no jobs end, the extra 0.11 job offers from applications with reports is sufficient to generate the treatment effect on employment. But any such exercise would require strong assumptions.

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. Only 22% of candidates are interviewed within 3 months of treatment. Hence the recall periods for the search effort questions will generally miss the initial two month period when employment is rising.[18]

4.2 Workseekers Face Information Frictions

In this section we explore whether revealing information only to workseekers can replicate the employment and earnings effects of our certification intervention. If not, then it is likely that certification does not work by addressing only workseeker-side information frictions.

We implement a 'private' certification intervention, distinct from the 'public' certification intervention described above. 2,114 candidates over 27 assessment days are randomized into a private certification group simultaneously with the public certification and control groups. The three groups are balanced on baseline characteristics (Table C.1). Candidates assigned to the private certification intervention receive an unbranded, anonymous certificate with the assessment results rather than the branded, identifiable 'public' certificate (Figure 3).

We interpret the private treatment as primarily providing information to the workseekers about their own types. Candidates in this group receive only one copy of the report, printed on low-quality paper, and do not receive an electronic version. Candidates can photocopy the private certificate and share it 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 identity number), use Harambee's name but not any branding, and are not branded by the World Bank. Candidates receive a briefing from a psychologist about the assessment results. But this briefing does not encourage them to share the certificate with firms or mention that this is possible, unlike the briefing received by candidates in the public certification group. Candidates in the public certification, private certification, and control groups all receive the same one hour of job search counselling and email with job search advice.

The private and public certification interventions have similar effects on workseekers' beliefs and search targeting. The private treatment makes workseekers' beliefs about their own skills more accurate (Table 3 column 1).[19] The private treatment has no effect on generalized self-esteem, like

[18] Consistent with this timing explanation, effects on all search effort measures are marginally larger for respondents with a shorter time between treatment and endline. This result is robust to instrumenting the treatment-to-endline time with the random order in which candidates were assigned to be surveyed.

[19] This effect is slightly smaller than the corresponding public treatment effects on beliefs. The private effect on beliefs about skills may be smaller because the public treatment conveys information differently (e.g. the branding makes it more credible to workseekers) or because the information is more likely to be retained (e.g. workseekers are more likely to still have copies of the public report or discuss it in recent job interviews). To separate these hypotheses, we measure workseekers' beliefs about their skills using a text message survey 2-3 days after treatment. The public and private effects in this survey are not different to each other, suggesting the difference in the endline survey 3-4 months later is due to differential retention. See Appendix C and Table C.9 for details.

Figure 3: 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.

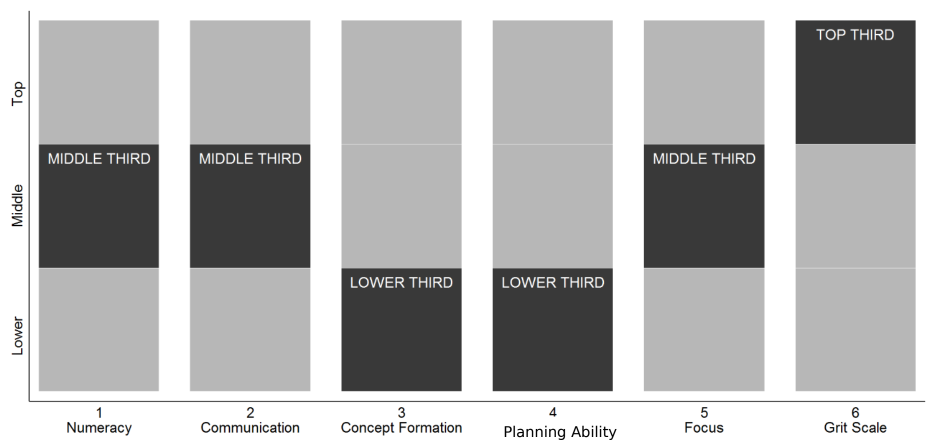
1. The Numeracy tests measure various maths abilities. Your score is the average of the two maths tests you did today at Harambee.
2. The Communication test measures English language ability through listening, reading and comprehension.
3. The Concept Formation test measures the ability to understand 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.

the public treatment (column 2). The private and public effects on search targeting are almost identical (column 3). Candidates in the private arm expect to receive 5% more offers than control candidates, suggesting that they view the new information and search targeting as useful (column 8).

However, the private certification intervention has smaller effects than public certification on candidates' outcomes in the labor market. Private certification effects on the probability of employment and hours worked are positive but small, not significantly different to zero, and significantly smaller than the public certification effects (columns 9-10). Private certification increases earnings and wages but both effects are less than half the size of the public certification effects and significantly smaller (columns 9-10).

These results are consistent with quantitatively important information frictions facing both workseekers and firms. Giving information to workseekers changes their beliefs about their skills, allowing better search targeting, and leading to higher earnings. When they can also credibly convey that information to firms, they are more likely to be employed and have even higher earnings and wages.

It is possible but unlikely that these results reflect only firm-side or only workseeker-side information frictions. First, the private certification may deliver some information to firms and this information, rather than changes in workseeker beliefs or search targeting, may generate the positive private effect on earnings. This explanation is consistent with the fact that candidates report using some private certificates in job applications, although private effects on certificate use are on average one third as large as the public effects (Table 3 columns 4-7). We cannot conclusively rule out this interpretation. But we view it as unlikely: the private certificates were designed not to be credible to firms, using feedback from informal interviews with hiring managers.

Second, the public certification effects on employment and earnings may be larger than the private effects because workseekers incorrectly believe that firms face information frictions. Under this explanation, workseekers in the public certification group believe that firms are more likely to respond to job application submitted with certificates, hence they search more or search differently. These changes in search behavior may generate higher employment and earnings, even if firms face no information frictions and do not respond to certificates. In the next section we address this possibility by discussing an experiment that directly manipulates firms' information, without any scope for changes in workseeker behavior. This explanation is difficult to reconcile with the zero public certification effects on multiple measures of search effort discussed in Section 4.1.

4.3 Firms Face Information Frictions

In this section we explore whether revealing information only to firms changes their responses to job applications, without allowing any potentially mediating behavior by workseekers. If so, it is likely that firms face information frictions and that the employment and earnings effects of certification are explained partly by firm-side information frictions.

We run an audit-style study to answer this question. We describe the experiment briefly here, with more details in Appendix D. We invite a random sample of assessed candidates to send us a resume that we forward to prospective employers on their behalf. We create a list of job vacancies by scraping online job advertisements. We eliminate scam vacancies and vacancies that require work experience or university education, where many candidates in our sample would be ineligible. We send resumes from 4 randomly chosen candidates to each vacancy, each from a different email address. We generate two outcome variables based on the email responses from firms. 'Interview invitations' are invitations to interview with the firm. 'Any responses' are similar to 'callbacks' in audit studies and include interview invitations and requests to provide more information by email or by visiting the firm in person.

We randomize each vacancy to receive either 1 or 3 resumes with public certificates attached. We also randomize which of the resumes are chosen to receive public certificates. This design motivates the estimating equation

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

where Y_{iv} is the response to resume r sent to vacancy i, $\text{Certificate}_{\text{iv}}$ is an indicator equal to one if the application includes a public certificate, $\text{HighIntensity}_{\text{v}}$ is an indicator equal to one if the vacancy receives 3 rather than 1 application with certificates, V_v is a vector of vacancy fixed effects that subsumes the main effect of $\text{HighIntensity}_{\text{v}}$, and E_{rv} is a vector of fixed effects for the email addresses used to submit the applications. We cluster standard errors by resume and vacancy.[20]

β_1, the effect of using a public certificate when other applications do not, is positive. Applications with a public certificate are 1.6 percentage points more likely to get a response and 1 percentage point more likely to get an interview invitation (Table 4). These are substantial effects, equal to respectively 13 and 11% of the mean response and interview rates, although the interview effect is only marginally statistically significant. Results are very similar when we remove the vacancy or email address fixed effects or include controls for resume-level characteristics.

These results show that more informative applications lead to higher callback and interview invitation rates in a low-information environment. This suggests a role for firm-side information frictions in explaining the earnings and employment effects of public certification. Combining this result with the observed effects of the public and private certification on workseekers' beliefs, search behavior, and outcomes in the labor market suggests that both firms and workseekers' face information frictions.

β_2, the change in the effect of using a public certificate when other applications also use a cer-

[20]Like most audit studies, we submit the same resume to multiple vacancies. Each resume includes a certificate for half of these vacancies. Audit studies generally cluster standard errors by resume (Neumark, 2018). Abadie et al. (2017) recommend clustering by the unit at which treatment is assigned. We therefore cluster by both vacancy and resume. Results are very similar when clustering only by vacancy or only by resume.

Table 4: Treatment Effects of Additional Information in Audit Study

	(1)	(2)	(3)	(4)
	Any response		Interview request	
Certificate (β_1)	0.016	0.017	0.009	0.010
	(0.009)	(0.010)	(0.004)	(0.006)
Certificate × HighIntensity (β_3)	-0.027	-0.027	-0.014	-0.015
	(0.014)	(0.014)	(0.009)	(0.010)
Outcome mean	0.127	0.127	0.085	0.085
# applications	3752	3752	3752	3752
Email address fixed effects	No	Yes	No	Yes
Vacancy fixed effects	No	Yes	No	Yes

Note: Coefficients are from regressing each outcome on a vector of treatment assignments. Heteroskedasticity-robust standard errors shown in parentheses clustered by resume and vacancy. The vacancy-level treatment is included in columns 1 and 3 but omitted in columns 2 and 4 because HighIntensity is colinear with the vacancy fixed effects.

tificate, is negative. Applications that include a public certificate are 2.7 percentage points less likely to get a response and 1.5 percentage points less likely to get an interview invitation when they compete against other applications with reports. This result suggests that more informative applications may not be more valuable in a higher-information environment. However, more informative applications may still be valuable for job offers, as opposed to callbacks and interviews, in a higher-information environment. If firms use callbacks and interviews to get more information, then certificates may allow them to interview fewer candidates and still make better-matched hires.[21]

There are, however, some caveats to the interpretation of the audit study results. This examines only one hiring method (online applications) and one stage of that process (interview invitations). These are standard limitations of audit studies. But it does mean that the design cannot easily quantify the importance of firm-side information frictions relative to workseeker-side information frictions on the same outcome. Furthermore, we randomly match workseekers to vacancies in the audit study. This omits any role for search targeting, which the public and private certification results suggest may be important. We therefore view the audit study as less important evidence than the arms of the workseeker-facing experiments. It simply provides additional evidence that the difference between the public and private certification results can be explained by firm-side information frictions.

5 What Do Workseekers and Firms Learn From Skill Certification?

The preceding two sections show that skill certification provides information that improves workseekers' labor market outcomes. In this section we explore what workseekers and firms learn from skill certification, what this implies for the effects of certification on different types of workseek-

[21] Jarosch and Pilossoph (2019) make a similar argument both theoretically and empirically about audit studies that manipulate firms' information about applicants' employment history.

ers, and what this might imply for general equilibrium effects of a richer information environment. This section relies on smaller experiments and heterogeneity analysis of the main experiments. We interpret these as more cautious extension results, relative to the more confident results from the preceding two sections that are the main focus of the paper.

5.1 Assessment Results Matter, Not Just Being Assessed

We conduct two smaller experiments that manipulate information about workseekers' assessment results, holding constant the information that workseekers have been assessed. The first experiment is a 'placebo' certification. We randomly assign 254 candidates from our workseeker sample, assessed over 3 days, to this treatment arm. These candidates receive placebo certificates that are identical to the public certificates in all ways but one - the actual assessment results are omitted (Figure E.1).[22]

The placebo certification treatment has small and statistically insignificant effects of labor market outcomes (Table E.1). It increases an index of labor market outcomes by only 0.03 standard deviations (standard error 0.04 standard deviations), compared to a public certification effect of 0.12 standard deviations. This index is an inverse covariance-weighted average of the five labor market outcomes discussed in Section 3.3: employment, hours, earnings, wages, and contract status. The placebo certification effects on the five individual outcomes are all smaller than the public certification effects and are on average only 26% as large. But we cannot reject equality of the public and placebo effects for all of the individual outcomes because the small size of the placebo sample leads to large standard errors.

The second experiment measures firms' willingness-to pay (WTP) for information on workseekers' assessment results. We recruit 69 establishments located in commercial areas near the low-income residential areas in Johannesburg where most workseekers in our sample live and are likely to work.[23] We conduct a survey and WTP exercise with the person responsible for hiring decisions at each of these establishments. We show this person a secure online database containing assessment results, contact information, and selected resume-style information for our 6,891 candidates. This database allows firms to filter and search for candidates with specific types and obtain their contact information. See Figures F.1 and F.2 for selected screenshots of the database. We

[22] The placebo certificates contain the same branding and logos as the public certificates, the same identifying information about candidates, and the same information about Harambee and the assessment process. Candidates in the placebo arm receive an email copy of their certificate and 20 physical copies printed in color on high-quality paper. They receive a briefing with psychologists covering the same information on how to use the certificates in job search as the public certification in candidates, except that the briefing does not discuss the assessment results.

[23] We recruit establishments by asking 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 establishments that have hiring responsibilities. This includes some single-establishment firms and some branches of larger firms where hiring decisions are made by branches. Most firms are in retail, have multiple entry-level workers, expect to hire entry-level workers in the next year, and take on average four weeks to fill a vacancy (Table F.1). These firms' location and sector are similar to those where our candidates are likely to apply for jobs.

measure WTP using a Becker-DeGroot-Marschak mechanism. We tell firms the database normally costs 10,000 South African Rands (USD 1,600 PPP) for three months access, ask how much they are willing to pay for access, and then randomly offer them a discount between 0 and 100%. If their stated WTP is higher than the normal price minus the discount, we give them access to the database. If their stated WTP 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' WTP is substantial: 68% of firms report positive WTP and the unconditional mean WTP is 1,161 South African Rands or USD 186 PPP (Figure F.3). This mean WTP is 224% of the mean weekly earnings for employed candidates in our workseeker sample. This measures WTP for the database of with assessment results relative to the placebo database. Firms know all candidates in the placebo database have been assessed, so this WTP captures the marginal value of learning assessment results to firms.

Both the placebo certification intervention and WTP measurement show that information about assessment results is valuable. This shows that the public treatment effects are not explained by firms simply learning that workseekers have been assessed, which might be interpreted as a signal of tenacity or proactivity. The placebo certification results also show that public certification does not change labor market outcomes simply by making applications look more professional or grab more attention. This provides additional support for our preferred interpretation: that public certification provides information about workseekers' types and facilitates more and more productive firm-worker matches.

5.2 Certification Facilitates Horizontal More Than Vertical Differentiation

Our conceptual framework distinguishes two types of differentiation. Under vertical differentiation, type i workseekers are more productive than type j workseekers in type i and j jobs. Under horizontal differentiation, type i workseekers are more productive than type j workseekers in type i jobs and vice versa. In this section we discuss three relevant patterns in our data, two consistent with horizontal differentiation and one inconsistent with vertical differentiation.

First, there is substantial heterogeneity in firms' relative demand for different skills. We show this using an incentivized choice experiment with the sample of 69 establishments described in the previous subsection. We ask the person at each establishment responsible for hiring to rank profiles of seven hypothetical candidates and tell them we will use their ranking to match them with workseekers from the online database, following Kessler et al. (2019). Six of the profiles have middle terciles for five assessments, and a top tercile for one assessment. There is substantial variation in firms' relative ranking of profiles (Table F.2). All profiles' median rank is between second and fourth. The share of firms ranking each profile highest ranges from 6 to 33%. The share of firms ranking each profile lowest differs by at most 9 percentage points. All profiles' median rank is between second and fourth. The seventh profile has middle terciles for all six assessments and has a one-year post-secondary certificate, while the other six profiles have only completed secondary school. Only 9% of firms rank this profile first and 76% of firms rank this last, showing that firms value the assessed skills relative to an alternative signal of productivity in which workseekers might invest.[24]

Second, assessment results are weakly correlated across skills within candidate. Numeracy and concept formation are most highly correlated with $\rho \approx 0.5$. But most other pairwise correlations are substantially lower, with $\rho < 0.1$ for some pairs of skills (Table C.3). As a result, most candidates' certificates show substantial variation across skills. 88% of the candidates have at least one top tercile but only 24% of candidates have four or more top terciles and only 2.3% of candidates have all top terciles. 76% of the candidates have at least one bottom tercile but only 12% of candidates have four or more bottom terciles and only 0.7% of candidates have all bottom terciles. 64% of candidates have both top and bottom terciles. This pattern is not unusual, with weak correlations across skills within candidate in other studies that measure multidimensional skills.

Third, we do not find strong evidence of vertical differentiation in the public certification experiment. To test for vertical differentiation, we construct three single indices that combine the multidimensional assessment results in different ways: the number of top terciles minus bottom terciles, the first principal component of the cardinal scores, and a weighted average of the cardinal scores with weights based on their association with earnings.[25] The first index weights all skills equally, the second gives more weight to skills that are highly correlated with each other, and the third gives more weight to skills with higher associations with earnings. For each index, we construct an indicator for above-median values of the index. We then augment equation (1) to include this indicator the indicator interacted with treatment assignments. The interaction effects on employment are smaller than 2 percentage points and not significantly different to zero for all indices (Table C.10).[26]

Taken together, these results are more consistent with horizontal than vertical differentiation. Firms and workseekers seem to face information frictions, and certification provides information

[24]We conduct a second experiment where we ask firms to rank profiles with assessment results shown for some skills and concealed for others. This assess whether firms value information about specific skills as well as the level of the skills. The two experiments may yield different results if, for example, firms find skill S_1 most valuable but believe the assessments of skill S_2 yield more new information. This second experiment also shows substantial heterogeneity in firms' ranking of different profiles.

[25]The weights equal the coefficients from regressing earnings on the cardinal scores using control group data. This index assigns most weight to the communication score. Results are similar for weighted averages based on the coefficients of regressions of control group earnings on polynomial or spline functions of the skills.

[26]Results are similar using the continuous indices instead of binary indicators, but the binary indicators make comparison of magnitudes across indices easier. We see no strong evidence of interaction effects between treatment and higher scores using alternative model specifications: allowing nonlinear interactions between skill indices, using different single indices, or using machine learning methods to estimate heterogeneous treatment effects simultaneously across all individual scores.

that facilitates more productive firm-worker matches that generate higher employment and earnings. Employment and earnings rise for many different types of workers, not just those with high values of a skill index of the six assessment scores. This is consistent with models of multidimensional skill where information frictions can lead to poor matches between workseeker skills and firm requirements (Lise and Postel-Vinay, 2020).

However, our experiment is not designed to test horizontal versus vertical differentiation, and certification may facilitate vertical differentiation in different samples or with different assessments. Certification is more likely to facilitate vertical differentiation if it assesses one skill, assesses multiple highly correlated skills, or reports a single summary measure of multiple skills. For example, Pallais (2014) studies the effect of publicizing single-dimensional evaluations of workers' past performance on their future labor market outcomes. Our sample by design excludes workseekers with less than complete high school education or with university education. There may be substantial vertical differentiation in the full labor market, with mainly horizontal differentiation within the relatively homogeneous sample of workseekers we study.

5.3 Certification Is More Effective When Other Information on Workseekers' Skills is Limited

If certification changes labor market outcomes by providing information about workseekers' skills, then it should be most effective when there are limited alternative sources of information on workseekers' skills. These sources might include past work experience and post-secondary education, which allow workseekers and firms to learn about workseekers' productivity in specific tasks. We test this idea by augmenting equation (1) to include interactions between treatment and proxies for alternative sources of information. Treatment effects on employment 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 C.10). We also estimate the latent probability of being employed at endline as a single summary measure.[27] 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 can substitute for traditional sources of information about workseekers' skills.[28] This is consistent with evidence that educational qualifications are more useful for members

[27]We estimate the latent probabilities following Abadie et al. (2018). We regress endline employment 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. Baseline employment is the most important predictor of endline employment.

[28]This result is not explained by a correlation between workseekers' skills, education, and past employment. We regress employment on treatment assignment, a single index measure of skill from Section 5.2, a measure of information about workseekers' skills from this section, and interactions between the latter two measures and treatment assignment. The interactions between treatment and the single index skill measure remain close to zero, while the interactions between treatment and the measure of information about workseekers' skills remain negative.

of groups facing statistical discrimination (Arcidiaono et al., 2010).

5.4 Possible General Equilibrium Effects of More Information

We show that small-scale use of certifications to reduce information frictions increases employment for certified workseekers. Our experiment is not designed to speak to identify the general equilibrium effects of market-level increases in information about workseekers' skills. In this section we briefly discuss some features of our results that may speak to general equilibrium considerations.[29]

Our conceptual framework shows how providing more information through certification can have substantial effects at both small and large scales. More information allows workseekers to apply to jobs where they will be more productive and firms to hire workseekers who will be more productive in the jobs they offer. This allows higher output and earnings from each match and increases the share of latent matches that generate enough value to pay above minimum or reservation wages. This interpretation matches many features of our core results: certified workseekers update beliefs about their skills, target their search differently, obtain more job offers, are more likely to be employed, and earn more conditional on being employed. This interpretation also matches features of the extension results discussed in this section: firms have heterogeneous preferences for skills, firms value learning specific assessment results, and hence gains from certification are not limited to a few workseekers with specific skill profiles.

There are alternative frameworks that predict substantial certification effects at small scale and zero effects at large scale. Consider a modified version of our conceptual framework where jobs pay heterogeneous wages (perhaps due to heterogeneous firm production technology) and workseekers are homogeneous so each workseeker produces the same output in each job. Workers are homogeneous but wages vary across jobs, so more workseekers apply to high-wage jobs and firms are indifferent about which workseekers to hire. In this framework high-wage firms might use certificates a tie-breaker when deciding who to hire, explaining why public certification increases workseekers' employment and earnings and why public certificates increase callbacks in the audit study but only when few applications use certificates.

However, this alternative framework does explain why workseekers update their beliefs about their skills and search differently when they learn their assessment results, or why workseekers earn more even when they cannot easily convey their assessment results to firms. It also does not explain why firms are willing to pay for access to assessment results and why the placebo certificates that do not show assessment results do not improve workseekers labor market outcomes.

Even if reducing information frictions at large scale has no effect on employment, it may still raise workseeker or firm welfare by reducing job search and vacancy posting costs and reducing the

[29] Comparisons of effects of specific active labor market policies running at smaller and larger scales is limited and evidence is mixed. 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.

frequency of bad hires that lead to separations (Donovan et al., 2018).

6 Conclusion

Firms make hiring decisions and workseekers make job search decisions based on potentially noisy information about workseekers' skills. Providing more information about workseekers' skills may reduce these information frictions and hence improve workseekers' outcomes in the labor market. We show that assessing workseekers' skills and communicating the assessment results to both workseekers and firms increases employment by 17% (5 percentage points), earnings by 34%, and hourly wages by 20% for the assessed workseekers. These results show that certification gets more workseekers into work and gets workseekers into higher-paying jobs. Additional experiments show that both workseekers and firms face information frictions. The distinction between workseeker- and firm-side frictions is important, as it informs how government or private firms might design information-provision products.

We study a context and sample where information frictions are likely due to limited work experience and weak education-skill relationships, hiring mistakes are costly, and reservation and minimum wages are relevant. However, none of these features are unique to young workseekers in South Africa. Formal education qualifications are weakly related to measured skills in many countries (Pritchett, 2013). Many labor markets face more regulations governing hiring, firing, and probation than in South Africa (Botero et al., 2004). Hiring mistakes may be costly even when separations are unregulated, due to reposting and retraining costs. High rates of youth unemployment in many countries are consistent with information frictions, as youths have less job search and work experience that can reveal their skills to themselves or to firms (International Labour Organization, 2017).

Our results suggest there may be scope for market-based provision of information about workseekers' skills. We show that firms are willing to pay for access a database with information on workseekers' skill assessment results and contact information. We also ask workseekers in our sample how much of a hypothetical job search subsidy they would be willing to spend on certification. They report 17%, compared to 24% on training and 27% on transport, suggesting the possibility of charging workseekers for assessment services. Some large firms already use some in-house psychometric assessments in hiring (Autor and Scarborough, 2008; Hoffman et al., 2018). Anecdotally, psychometric assessments seem rarer in small firms, perhaps because in-house assessment systems are unlikely to be cost-effective when hires are infrequent. There are some third-party providers of assessment services around the world, including Harambee, LinkedIn, and the Manpower Group. Our results suggest that providing more information through certification can be valuable even when some firms already use assessments, suggesting there may be scope to grow this market. There are important market design questions around third-party provision that might be addressed in future work, such as which side(s) of the market will pay for assessment services, how third-party providers might establish reputations, and under what conditions participants will opt into or out of assessment.

Our results also motivate future work on the interaction between different information provision mechanisms. For example, we find that public certification is most effective for workseekers with less work experience and without university education. This suggests that skill assessment and certification can substitute for alternative sources of information about workseekers' skills. Future work could examine conditions under which skill assessment and certification are complements or substitutes for network referrals, reference letters, or outsourcing agencies.[30]

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 (2019): "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City," *Review of Economic Studies*, forthcoming.

Abel, M. (2019): "Unintended Labor Supply Effects of Cash Transfer Programs: New Evidence from South Africa's Pension," *Journal of African Economies*, 28, 558–581.

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

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

Aigner, D. and G. Cain (1977): "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 30, 175–187.

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.

[30] We one result consistent with certification enhancing the effectiveness of referrals, potentially by helping network links to target referrals or making their referrals more credible to employers. Public certification slightly increases 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. However, this result is only marginally statistically significant once we account for multiple testing across the different ways of finding a job. Hence we view this as a suggestion for future work, rather than a core result.

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 (2019): "Screening and Signaling Non-Cognitive Skills: Experimental Evidence from Uganda," Manuscript, University of Southern California.

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.

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.

Bowling, A. (2014): *Research Methods in Health: Investigating Health and Health Services*, Maidenhead, GB: McGraw Hill; Open University Press, 4 ed.

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.

Card, D., J. Kluve, and A. Weber (2018): "What Works? A Meta-Analysis of Recent Active Labor Market Program Evaluations," *Journal of the European Economic Association*, 16, 894–931.

Chamorro-Premuzic, T. and A. Furnham (2010): *The Psychology of Personnel Selection*, Cambridge University Press.

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

Conlon, J., L. Pilossoph, M. Wiswall, and B. Zafar (2018): "Labor Market Search with Imperfect Information and Learning," Working paper 24988, National Bureau of Economic Research.

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.

Cronbach, L. J. (1951): "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16, 297–334.

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.

Doppelt, R. (2016): "The Hazards of Unemployment: A Macroeconomic Model of Job Search and Resume Dynamics," Working Paper, Pennsylvania State University.

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.

Falk, A., D. Huffman, and U. Sunde (2006): "Self-Confidence and Search," IZA Discussion Paper 2525.

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.

Gonzalez, F. and S. Shi (2010): "An Equilibrium Theory of Learning, Search, and Wages," *Econometrica*, 78, 509–537.

Gronau, R. (1974): "Wage Comparisons – A Selectivity Bias," *Journal of Political Economy*, 82, 1119–1143.

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.

Heckman, J. (1974): "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679–694.

Heckman, J., R. LaLonde, and J. Smith (1999): "The economics and econometrics of active labor market programs," in *Handbook of Labor Economics*, Elsevier, vol. 3, 1865–2097.

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.

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

International Labour Organization (2017): *Global Employment Trends for Youth 2017: Paths to a better Working Future*, Geneva: International Labour Office.

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.

Jarosch, G. and L. Pilossoph (2019): "Statistical Discrimination and Duration Dependence in the Job Finding Rate," *Review of Economic Studies*, 86, 1631–1665.

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.

Jovanovic, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87, 972–990.

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.

KENNAN, J. (2010): "Private Information, Wage Bargaining and Employment Fluctuations," *Review of Economic Studies*, 77, 633–664.

KERR, A. (2017): "Tax(i)ing The Poor? Commuting Costs in South African Cities," *South African Journal of Economics*, 85, 321–340.

KESSLER, J., C. LOW, AND C. SULLIVAN (2019): "Incentivized Resume Rating: Eliciting Employer Preferences without Deception," *American Economic Review*, forthcoming.

KLUVE, J., S. PUERTO, D. ROBALINO, J. M. ROMERO, F. ROTHER, J. STÖTERAU, F. WEIDENKAFF, AND M. WITTE (2019): "Do youth employment programs improve labor market outcomes? A quantitative review," *World Development*, 114, 237–253.

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.

LAWRENCE, I. AND K. LIN (1989): "A Concordance Correlation Coefficient to Evaluate Reproducibility," *Biometrics*, 45, 255–268.

LEE, D. (2009): "Trimming, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071–1102.

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. AND F. POSTEL-VINAY (2020): "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, forthcoming.

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.

MANSKI, C. (1989): "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343–360.

MCCALL, B., J. SMITH, AND C. WUNSCH (2016): "Government-sponsored vocational education for adults," in *Handbook of the Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 5, 479–652.

MUELLER, A., J. SPINNEWIJN, AND G. TOPA (2018): "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence and Bias," Working paper 25294, National Bureau of Economic Research.

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.

POWELL, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

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.

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.

SPINNEWIJN, J. (2015): "Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs," *Journal of the European Economic Association*, 13, 130–167.

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. (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.

A Further Details on Assessments

A.1 Assessments

We assess each workseeker's skills in six domains. Detailed information on all six assessments is available at https://www.assessmentreport.info/, including sample questions. Most of the assessments are already used by Harambee and by some large firms in South African during hiring. We do not claim that these are best possible assessments for predicting workplace performance. But these are assessments that some market agents have chosen to use, have reasonable psychometric properties, and are correlated with workplan performance in some existing research.

All assessments are conducted in English, the same language used for all Harambee interaction with candidates. All assessments are conducted on desktop computers, so the assessment results may be sensitive to candidates' computer skills. To minimize this sensitivity, all candidates do some practice computer exercises before the assessments and all 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.

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. It is similar to the Raven's Progressive Coloured Matrices assessment (Raven, J. and Raven, J., 2003). In the South African context, it is correlated with workplace performance measures for clerks working at a large city and several private firms, within a large financial institution, and during training for new employees at a large financial institution (Lopes et al., 2001; De Kock and Schlechter, 2009; Taylor, 2013). The very similar Raven's test is widely used in hiring and selection (Chamorro-Premuzic and Furnham, 2010).

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, as they believe it identifies some of the skills needed by cashiers. The simpler assessment was developed by a South African adult education provider (https://www.mediaworks.co.za/) and is designed to assess comfort with arithmetic used in high school. 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.

Communication captures English language listening, reading and comprehension skills by testing comprehension of spoken and written passages. The assessment was developed by a South African adult education provider (https://www.mediaworks.co.za/) and is designed to assess English proficiency for high school students.

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 moderate the negative effects of workplace related stress such as burnout and absenteeism in service sector jobs (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 predict retention rates among truckers in the US conditional on cognitive skills (Burks et al., 2009).

For the first 17 of the 84 assessment days, covering 26% of candidates, we used self-reported measures of *control* and *flexibility* instead of the focus and planning tasks. 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.

None of the main results in the paper are substantially different between the sample using the focus and planning assessments and the sample using the control and flexibility assessments. The assessment scores are used in the paper in three ways. First, we use assessment scores as a prespecified conditioning variable when estimating treatment effects. We use the concept formation, communication, grit, and numeracy scores individually for this purpose. We combine the remaining scores into a single measure by taking the first principal component of control and flexibility and standardizing it, taking the first principal component of focus and planning and standardizing it, and then appending the two principal components together. Second, we use assessment scores in the heterogeneity analysis described in Section 5.2. We use only the scores observed for all candidates (concept formation, communication, grit, and numeracy) for this analysis. Results are generally similar when we restrict to the 74% of candidates who took the focus and planning assessments and use all six assessments. Third, we use assessments in the firm-facing experiments described in Sections 5.1 and 5.2. The online platform reports all eight assessment results and explains why two results are missing for all candidates. The profile-ranking exercise does not use the control or flexibility scales.

We conduct some psychometric validation to verify that the grit, control, and flexibility assessments are usable in research, following Esopo et al. (2018). Harambee's assessment infrastructure does not allow us to observe item-level responses for the concept formation, communication, or numeracy assessments. First, we conducted cognitive debriefings with 20 Harambee candidates. 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.

Second, we estimated the extent to which different items in each assessment move together, using Cronbach's alpha (Cronbach, 1951). All assessments have $\alpha > 0.65$. Values of 0.6 or 0.7 are sometimes used as minimum thresholds for psychometric assessments. Third, we administered the assessments twice for 150 candidates, with ten days between assessments. We estimated Lin's Concordance Correlation Coefficient (Lawrence and Lin, 1989) between the two test observations. All assessments have $\rho_c > 0.62$. Fourth, we check if any items on the scales have very low variation across candidates using the maximum endorsement frequencies. No items meet the threshold for being dropped due to insufficient variation from Bowling (2014).

All assessments are 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.

The terciles shown on the assessment results 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 the other skills. Tercile assignments are largely unchanged if we retrospectively construct them using our full sample of assessed workseekekers.

A.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 helped over 20,000 candidates secure entry-level jobs using these assessments 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 assessment results they request from Harambee. All firms used at least one assessment to screen candidates and 73% of firms used all three assessments. In contrast, only 57% required a high school graduation certificate with results on the standardized graduation exam, and 3% required references. This suggests firms find this skill information useful relative to other traditional sources of information about prospective workers' productivity. Harambee also administers a set of career aptitude measures provided by a psychometric testing firm. 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 the certification because it is a proprietary instrument whose psychometric properties we could not assess.

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

		% of firms using each piece of information to screen candidates						
		Assessment result for			Career	Criminal	High school	
Sector	# firms	Communi-cation	Concept formation	Numeracy	aptitude profile	record check	graduation certificate	Reference
Hospitality	11	0.82	1.00	0.91	0.64	0.91	0.64	0.00
Retail	16	0.69	0.56	0.88	0.81	0.94	0.75	0.06
Corporate	6	1.00	1.00	0.83	0.33	1.00	0.00	0.00
Total	33	0.79	0.91	0.79	0.67	0.94	0.58	0.03

Table shows use of assessment results and other information by 33 firms that have long-term recruiting relationships with Harambee. 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. The criminal background check is a set of checks against government records that the candidate had no criminal record or bad credit history. We observe only what information these 33 firms request from candidates, not how they use the information once they get it.

B Labor Market Effects at the 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 the 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:

$$\begin{aligned}
&\underbrace{\mathbb{E}[Earn|Treat=1]-\mathbb{E}[Earn|Treat=0]}_{\text{ATE for earnings}} \qquad (3)\\
=\;&\underbrace{(\mathbb{E}[Earn|Treat=1,Work=1]-\mathbb{E}[Earn|Treat=0,Work=1])}_{\text{ATE for earnings}\,|\,\text{employment}}\cdot\underbrace{Pr[Work=1|Treat=1]}_{\text{Treated employment rate}}\\
+\;&\underbrace{\mathbb{E}[Earn|Treat=0,Work=1]}_{\text{Control earnings}\,|\,\text{employment}}\cdot\underbrace{(Pr[Work=1|Treat=1]-Pr[Work=1|Treat=0])}_{\text{ATE for employment}}.
\end{aligned}$$

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.[31]

All terms in equation (3) 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 (3). We obtain standard errors by estimating all quantities as a system and using the Delta method.

This decomposition applies to *realized* 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. Alternative methods are available for studying latent earnings. One set of approaches point identifies the average treatment effect on latent earnings by by modeling the

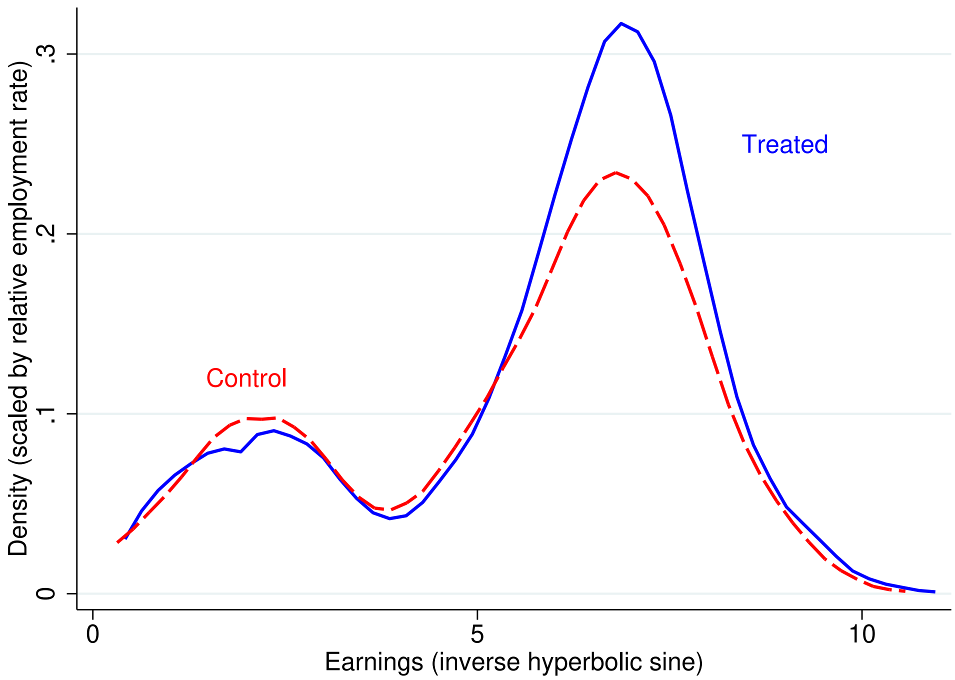
[31] Attanasio et al. (2011) show that the intensive margin effect can be further decomposed into two terms: the treatment effect on earnings conditional on candidates' baseline characteristics, and the difference in baseline characteristics between employed candidates in the treatment and control groups.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

selection process into employment and adjusting observed earnings for selection (e.g. Gronau, 1974 and Heckman, 1974). Another set of approaches bounds the average treatment effect on latent earnings by assuming that the earnings for the non-employed fall in some region of the observed earnings distribution (e.g. Lee, 2009 and Manski, 1989). Neither approach is ideal is our setting: the former methods require an instrument for selection into employment that we do not have and the latter methods will yield wide bounds given the large effect of public certification on employment. Another set of approaches point identifies quantile treatment effects on latent earnings by assuming that the earnings for the non-employed fall in some region of the observed earnings distribution (e.g. Powell, 1984). Our analysis of quantile treatment effects has a similar flavor to this approach, though we do not directly interpret these as effects on latent earnings.

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

The intensive-margin effect on earnings is also visible in the distributions and densities of earnings for the public certification and control groups. Figure 2 shows the distributions of earnings for each group and the quantile treatment effects of public certification. Figure B.1 shows the the densities of earnings for employed candidates in the control and treatment groups. We rescale the latter density by the ratio of treatment group to control group employment. Hence vertical differences between the densities represent treatment effects on the earnings densities unconditional on employment. The treatment effect on the earnings density is almost entirely above median earnings for employed control group candidates (6 IHS points, or 33 USD PPP per week). This shows that either the marginal candidates employed after treatment earn more than inframarginal control candidates or treatment increases earnings for inframarginal candidates.

Figure B.1: Density of Earnings in Control and Public Certification Groups



This figure shows the sample densities of earnings in the control and public cerification groups. To account for the positive treatment effect on employment, the treatment density is scaled by the ratio of employment in the treatment group to employment in the control group. The density is estimated only for the employed, so candidates with zero earnings are excluded.

C Additional Results Discussed in Paper

C.1 Summary Statistics and Balance Tests

This section reports summary statistics for the baseline workseeker sample (Table C.1) and endline workseeker sample (Table C.2). Balance tests for equal means of baseline measures are also reported in the final column of Table C.1. Table C.3 shows the correlation of assessment results for the different skills. Table C.4 compares age, gender, education, employment, and job search in our workseeker sample to the broader population of the Gauteng province where the study took place.

Table C.1: 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	0.583
Male	6891	0.382	0.486			0.267
University degree / diploma	6891	0.167	0.373			0.889
Any other post-secondary qualification	6891	0.212	0.409			0.642
Completed secondary education only	6891	0.610	0.488			0.794
Panel B: Assessment Results						
Numeracy score	6891	0.052	0.988	-1.187	1.411	0.523
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	0.764
Grit score	6891	0.031	0.992	-1.313	1.279	0.089
Other scores	6891	-0.002	1.070	-1.305	1.318	0.859
Panel C: Labor Market Measures						
Employed	6891	0.378	0.485			0.468
Earnings	2116	565	740	100	1400	0.083
Ever worked	6877	0.704	0.457			0.418
Panel D: Job Search Measures						
Searched	6891	0.968	0.175			0.058
Applications submitted[a]	6815	9.9	18.6	2.0	20.0	0.809
Search cost	6147	242	1520	30	400	0.276
Search hours	6699	17.0	20.8	2.0	48.0	0.231
Offers received[a]	6810	1.20	7.20	0.00	2.00	0.280
Panel E: Belief Measures						
Planned applications[a]	6840	48.9	1629.9	4.0	36.0	0.252
Fraction of assessments overconfident	6875	0.503	0.352			0.584
Fraction of assessments underconfident	6875	0.115	0.208			0.367

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 ≈ USD0.16 in purchasing power parity terms. Intensive-margin labor market measures are set to missing for non-workers. Intensive-margin search measures are set to missing for non-searchers. All assessment results are standardized to have mean zero and standard deviation one in the control group. Missing values reflect item non-response, mostly due to respondents reporting that they don't know the answer. All period-specific outcomes use a 7-day recall/forecast period unless marked with [a] (30-day recall/forecast period) or [b] (since treatment). The final column reports the p-value for testing equality of means of the baseline variables across all treatment groups, using heteroskedasticity-robust standard errors clustered by treatment date.

Table C.2: Summary Statistics for Endline Variables

Variable	# obs	Mean	Std dev.	10^{th} pctile	90^{th} pctile
Panel A: Labor Market Measures					
Employed	6607	0.323	0.468		
Earnings	2112	623	1183	2	1500
Hours worked	2121	28.5	21.6	4.0	56.0
Hourly wage	2097	33.1	72.3	0.1	77.8
Wage employment	2102	0.885	0.319		
Self employment	2102	0.114	0.318		
Panel B: Job Search Measures					
Any search	6608	0.692	0.462		
Applications submitted[a]	6577	12.8	21.5	1.0	27.0
Hours searched	6601	9.9	14.2	0.0	25.0
Search cost	6599	116	167	0	300
Responses[a]	6593	0.861	2.147	0.000	2.000
Offers[a]	6592	0.207	0.680	0.000	1.000
Panel C: Belief Measures					
Fraction of assessments overconfident	6607	0.345	0.237		
Fraction of assessments underconfident	6607	0.176	0.166		
Targeted search	6891	0.175	0.380		
Planned applications[a]	6591	16.1	29.7	3.0	30.0
Expected offers[a]	6531	4.49	5.70	1.00	10.00

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 labor market measures are set to missing for non-workers. Intensive-margin search measures are set to zero for non-searchers. Missing values reflect item non-response, mostly due to respondents reporting that they don't know the answer. All period-specific outcomes use a 7-day recall/forecast period unless marked with [a] (30-day recall/forecast period) or [b] (since treatment).

Table C.3: Correlations of Assessment Results

Panel A: Correlations In First 17 Days of Assessment (1,615 workseekers)					
	Concept formation	Grit	Numeracy	Control	Flexibility
Communication	0.337	0.127	0.386	0.237	0.126
Concept formation		0.108	0.489	0.174	0.098
Grit			0.163	0.507	0.334
Numeracy				0.212	0.107
Control					0.173
Panel B: Correlations In Remaining 67 Days of Assessment (5,276 workseekers)					
	Concept formation	Grit	Numeracy	Focus	Planning
Communication	0.346	0.088	0.393	0.171	0.258
Concept formation		0.094	0.519	0.225	0.292
Grit			0.129	0.049	0.106
Numeracy				0.162	0.325
Focus					0.181

Table shows pairwise correlation coefficients between assessment results. The sample is split because two of the assessments changed after the first 17 days of assessment, from the control and flexibility scales to the focus and planning tasks. See Appendix A for more details on the change. None of the pairwise correlations between the four assessments used for the entire period (communication, concept formation, grit, and numeracy) are substantively or statistically significantly different between the two periods.

Table C.4: Comparison between the Workseeker Sample and External Populations

	(1) South Africa	(2) Gauteng province	(3) Gauteng age 18-29	(4) Study sample of workseekers
Age	28.914	31.461	23.776	23.646
	(19.606)	(19.099)	(3.326)	(3.299)
Male	0.489	0.504	0.512	0.382
Currently Employed	0.290	0.375	0.346	0.378
Currently searching	0.102	0.151	0.302	0.984
< complete secondary school	0.737	0.612	0.429	0.011
Complete secondary school	0.184	0.255	0.449	0.610
> complete secondary school	0.072	0.120	0.116	0.379

Table compares the sample of workseekers in this study (column 4) to several external benchmarks: the country (column 1), the province of Gauteng where the study takes (column 2), and people in Gauteng in the eligible age range for the study (column 3). National and provincial statistics are calculated from the Quarterly Labour Force Surveys (QLFS), averaging over all 2016 waves. Calculations use post-stratification weights supplied by Statistics South Africa. QLFS data are not available by city but the greater Johannesburg metropolitan area where the study is conducted accounts for over half the population of the Gauteng province. Standard deviations are shown in parentheses for all continuous variables.

C.2 Benchmarking The Magnitude of The Earnings Effects

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

Minimum wage: During our study period, minimum wages in South Africa varied by sector and location. Sector- and location-specific minimum wages were either set by the Ministry of Labour or in bargaining councils, where large firms and unions agreed minimum wages that applied to all firms (Budlender et al., 2015; Isaacs, 2016). Table C.5 shows minimum wages for urban areas at the time of the study for several sectors relevant to workseekers in our sample.

Poverty Lines: South African poverty research often uses the cost of purchasing 2100 calories plus the average amount spent on non-food items by households whose food expenditure equals the food poverty line (Budlender et al., 2015; Leibbrandt et al., 2012). Using this definition, the adult monthly poverty line just before the study period was 1,386 South African Rands or USD 222 in purchasing power parity terms (Isaacs, 2016, p.22).

The average treatment effect on earnings is equal to 17% of the adult monthly poverty line or 6-10% of the monthly minimum wage at the time of the study.

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

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

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

Panel B: Benchmarking sample earnings and 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.

C.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 C.6). 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 C.6: Non-response by Treatment Group in Each Post-Treatment Survey Round

	(1) Text Message Survey	(2) Endline Phone Survey
Control	0.170	0.040
	(0.013)	(0.006)
Public	0.177	0.039
	(0.011)	(0.004)
Private	0.182	0.044
	(0.010)	(0.004)
Placebo	0.142	0.047
	(0.032)	(0.026)
p: Control = Pvt.	0.481	0.632
p: Control = Pub.	0.670	0.855
p: Pvt. = Pub.	0.785	0.388
p: Control = Pvt. = Pub.	0.778	0.681
p: Control = Plc.	0.414	0.787
p: Pvt. = Plc.	0.238	0.888
p: Pub. = Plc.	0.297	0.746
p: Control = Pvt. = Pub. = Plc.	0.641	0.841
# observations	6891	6891
# clusters	84	84

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

	(1) Text Message Survey	(2) Endline Phone Survey
Completed at most high school	-0.008	-0.003
	(0.012)	(0.005)
Numeracy score	-0.029	0.003
	(0.006)	(0.003)
Communication score	0.008	0.003
	(0.006)	(0.003)
Concept formation score	-0.019	0.002
	(0.006)	(0.003)
Grit score	-0.001	-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)	(0.005)
Perceived concept formation score	0.010	-0.003
	(0.009)	(0.004)
Self-esteem index	0.006	0.002
	(0.004)	(0.002)
Age	-0.002	0.001
	(0.001)	(0.001)
Male	0.049	0.014
	(0.010)	(0.005)
Employed	-0.005	-0.001
	(0.008)	(0.005)
Above median discount factor	0.012	0.007
	(0.009)	(0.005)
Respondent is present-biased	0.016	0.007
	(0.011)	(0.006)
Above median risk aversion	-0.007	0.001
	(0.008)	(0.005)
p: All coefficients jointly zero	0.000	0.041
Mean outcome	0.170	0.040
# observations	6891	6891
# clusters	84	84

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.

C.4 Additional Treatment Effects

Table C.8 shows the public certification effects of our main labor market outcomes without conditioning on the prespecified covariates. The results are very similar with or without the covariates.

Table C.9 shows public and private certification effects at two points in time: in the text message survey conducted 2-3 days after treatment and the endline phone survey conducted 3-4 months after treatment. This table shows four patterns. First, both treatments make candidates more likely to report that their assessment result matches their actual assessment result immediately after treatment. Second, both treatment effects decline over the following 3-4 months. Third, the public treatment effect on self-beliefs is significantly larger than the private effect after 3-4 months but not after 2-3 days. This suggests that the larger public treatment effect at 3-4 does not occur because the information it conveys is immediately more credible or easier to understand than the private treatment. Instead, it may be larger because the information is more memorable or the public treatment generates other effects, such as more job interviews or employment that provide more opportunities to learn about skills. Fourth, neither treatment affects generalized self-esteem at either point in time.

Table C.10 shows how treatment effects on employment vary by single index summary measures of candidates' skills (Panel A) and baseline candidate characteristics that might provide alternative measures of candidates' skills (Panel B).

Table C.11 reports public and private certification effects on all workseeker-level job search and labor market outcomes. These are organized into families of conceptually similar outcomes, which we use for multiple testing adjustments. First, we report q-values that control the false discovery rate across outcomes within each family (Benjamini et al., 2006). None of the q-values in this table are substantively different to the corresponding p-values reported in the main paper. 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 across all outcomes in the same family.

We omit some prespecified outcomes related to beliefs from this paper and analyze them in separate work. The search targeting measure discussed in Section 4 is not prespecified. We did not prespecify an analysis plan for the smaller extension experiments discussed in Section 5.

Table C.8: Treatment Effects on Labor Market Outcomes Without Covariates

	(1) Employed	(2) Hours[c]	(3) Earnings[c]	(4) Hourly wage[c]	(5) Written contract
Treatment	0.046	0.175	0.336	0.206	0.018
	(0.013)	(0.058)	(0.076)	(0.041)	(0.010)
Mean outcome	0.309	8.848	159.291	9.840	0.120
Mean outcome for employed		28.847	518.291	32.283	0.392
# observations	6607	6598	6589	6574	6575
# clusters	84	84	84	84	84

Coefficients are from regressing each outcome on a vector of treatment assignments and randomization block fixed effects without any other covariates. 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 [a] (30-day recall period) or [b] (since treatment). Outcomes marked with [c] use the inverse hyperbolic sine transformation. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

Table C.9: Treatment Effects on Self-Beliefs through Time

	(1)	(2)	(3)	(4)
	Perceived numeracy tercile 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

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. 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. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

Table C.10: Heterogeneous Treatment Effects on Employment

	(1)	(2)	(3)
Panel A: Heterogeneous Effects by Single Index Skill Measures			
Public treatment	0.052	0.052	0.054
	(0.011)	(0.011)	(0.012)
× Share top - share bottom terciles	0.019		
	(0.028)		
× PC_1(Scores)		0.004	
		(0.025)	
× Earnings-weighted average of scores			-0.007
			(0.029)
# observations	6607	6607	6603
# clusters	84	84	84
Panel B: Heterogeneous Effects by Alternative Information Sources			
Public treatment	0.052	0.052	0.051
	(0.011)	(0.012)	(0.012)
× post-secondary education	-0.027		
	(0.028)		
× employed at baseline		-0.043	
		(0.032)	
× $\hat{\Pr}$(Employed at endline |X)			-0.069
			(0.028)
# observations	6607	6607	6607
# 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 (measured skills, self-reported skills, education, age, gender, employment, discount rate, risk aversion). Heteroskedasticity-robust standard errors shown in parentheses, clustering by treatment date. The measures used for interactions in Panel A and column 3 of Panel B are indicators for above-median values of the underlying indices. All 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 control group employment status on the baseline covariates listed above and predicting employment for all candidates. Prediction for control group candidates uses leave-one-out-estimation to avoid overfitting. PC_1(Scores) is the first principal component of the skills. The earnings-weighted average of scores is the weighted average of the assessment results, with weights derived from a regression of control group earnings on assessment results.

Table C.11: Treatment Effects on Prespecified Outcomes with Multiple Testing Adjustments

	(1) Index	(2) Any search	(3) Applications[a,c]	(4) Search hours[c]	(5) Search cost[c]
Public	-0.012	-0.020	0.018	-0.035	-0.092
	(0.032)	(0.014)	(0.042)	(0.048)	(0.081)
Private	0.006	-0.006	0.036	-0.035	-0.031
	(0.032)	(0.014)	(0.038)	(0.049)	(0.088)
q: Public effect = 0		1.000	1.000	1.000	1.000
q: Private effect = 0		1.000	1.000	1.000	1.000
q: Public = private effect		1.000	1.000	1.000	1.000
Mean outcome	0.001	0.695	12.356	9.791	112.684
# observations	6608	6608	6577	6601	6599

	(1) Index	(2) Responses[a,c]	(3) Offers[a,c]	(4) Responses per application[a]	(5) Offers per application[a]
Public	0.016	0.023	0.006	0.000	-0.000
	(0.029)	(0.024)	(0.013)	(0.004)	(0.003)
Private	0.019	0.016	0.013	-0.005	0.001
	(0.026)	(0.022)	(0.013)	(0.004)	(0.004)
q: Public effect = 0		1.000	1.000	1.000	1.000
q: Private effect = 0		1.000	1.000	1.000	1.000
q: Public = private effect		1.000	1.000	1.000	1.000
Mean outcome	-0.023	0.871	0.195	0.099	0.030
# observations	6593	6593	6592	5944	5943

	(1) Index	(2) Used report[b]	(3) Applications with report[b,c]	(4) Interviews with report[b,c]	(5) Offers with report[b,c]
Public	NA	0.699	1.682	0.432	0.112
		(0.013)	(0.040)	(0.023)	(0.011)
Private	NA	0.289	0.572	0.144	0.036
		(0.012)	(0.033)	(0.017)	(0.008)
q: Public effect = 0		0.001	0.001	0.001	0.001
q: Private effect = 0		0.001	0.001	0.001	0.001
q: Public = private effect		0.001	0.001	0.001	0.001
Mean outcome		0.000	0.000	0.000	0.000
# observations		6609	6598	6597	6597

	(1) Index	(2) Employed in last week	(3) Employed in month 1	(4) Employed in month 2	(5) Hours[c]
Public	0.137	0.052	0.036	0.058	0.201
	(0.025)	(0.012)	(0.011)	(0.014)	(0.052)
Private	0.049	0.011	0.028	0.009	0.066
	(0.028)	(0.012)	(0.013)	(0.015)	(0.048)
q: Public effect = 0		0.001	0.001	0.001	0.001
q: Private effect = 0		0.504	0.142	0.504	0.336
q: Public = private effect		0.003	0.132	0.002	0.008
Mean outcome	0.001	0.309	0.465	0.437	8.848
# observations	6609	6607	6604	6607	6598

	(1) Index	(2) Earnings[c]	(3) Hourly wage[c]	(4) Written contract
Public	0.106	0.338	0.197	0.020
	(0.028)	(0.074)	(0.040)	(0.010)
Private	0.069	0.162	0.095	0.017
	(0.030)	(0.078)	(0.046)	(0.009)
q: Public effect = 0		0.001	0.001	0.019
q: Private effect = 0		0.066	0.066	0.066
q: Public = private effect		0.047	0.047	0.345
Mean outcome	0.006	159.291	9.840	0.120
# observations	6609	6589	6574	6575

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 the 84 treatment dates. Sharpened q-values control the false discovery rate across outcomes in each panel, following Benjamini et al. (2006). The first column of each panel shows inverse covariance-weighted averages of outcomes in each panel, following Anderson (2008). The index is omitted for the report use variables because these are zero for all control group candidates, so the covariance cannot be estimated. Mean outcome is for the control group. All outcomes use a 7-day recall period unless marked with [a] (30-day recall period) or [b] (since treatment). Outcomes marked with [c] use the inverse hyperbolic sine transformation. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

D Audit Study

To identify the effect of information provision on the firm side, we conduct an audit study. We submit real workseekers' applications to entry-level job vacancies and randomly vary the information firms see about workseekers' skills.We implement the audit study in eight sequential rounds (Appendix Table D.1).

Table D.1: 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

In each round, a subset of candidates who have completed the workseeker study endline is randomly selected. Selected candidates are invited by text message to submit application materials to us, within 7 days, for an undisclosed job opportunity.[32] 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. One "reminder" text message is sent to all candidates 1-3 days after this initial message.

Approximately 25% of the work seekers contacted across all rounds responded to our message within a week. Workseekers in the sample of the audit study are slightly selected and workseekers in the private certification group are overrepreseted relative to the full workseeker sample (Appendix Table D.2 Panel A). 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. The results are also robust to reweighting the audit sample to have the same distribution of treatment assignments and baseline covariates as the full workseeker sample.

Once candidates send their applications, they receive an automated acknowledgement. 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 (Appendix Table D.2 Panel B).

Simultaneously, entry-level job vacancies are identified from a number of online job posting sites. Selected vacancies are suitable for entry-level workers, such that all candidates in our sample would

[32]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.

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. Typical sectors include sales, admin, call center, industrial, restaurant, and service (Appendix Table D.3, Panel A).

Table D.2: Comparison Between Audit and Workseekers Study Samples

	Audit study sample			Workseeker sample		
	Mean	Std Dev.	Obs	Mean	Std Dev.	Obs
Panel A: 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
Panel B: 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	-	-	-

Table D.3: 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

For each workseeker 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.[33] We generate between 6 and 10 applications per work seeker in each round completed to date, so that the total number of applications equals 4 times the number of jobs. We do not represent ourselves as the candidate. Instead, a generic message and subject line are written for each of the four applications we submit to each vacancy, using four different email addresses.[34]

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 intensity. High treatment intensity vacancies get a public report in 3 of the 4 applications submitted. Low treatment intensity 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. At least one application received any response in 14 percent of the vacancies, and all applications received a response in 7 percent of the vacancies (Appendix Table D.3 Panel B).

We code application outcomes as missing if our application email bounces or receives a response that appears to be a scam. The resulting sample includes 3,752 applications from 632 candidates sent to 938 vacancies receiving 534 responses.

Appendix Table D.4 shows that the distribution of treatment assignments in the sample matches the design: half of a workseeker's applications are assigned to be control and half to the public report treatment; half the applications from each workseeker are sent to high saturation vacancies and half to low saturation vacancies. Of all applications submitted, only a small fraction (14 percent) receives

[33] 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 for candidate submissions to accumulate. However, job vacancies may become filled during that wait period.

[34] We send the following message: *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".*

Table D.4: 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

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). Rejections are uncommon and all results are robust to excluding these from the definition of "any response."

E Placebo Certification Experiment: Sample Report and Treatment Effects

Figure E.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.

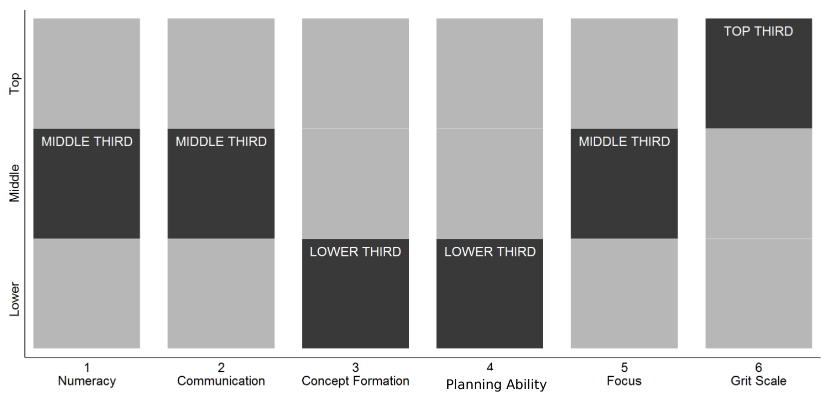
1. The Numeracy tests measure various maths abilities. Your score is the average of the two maths tests you did today at Harambee.
2. The Communication test measures English language ability through listening, reading and comprehension.
3. The Concept Formation test measures the ability to understand 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.

Table E.1: Public and Placebo Certification Effects

	(1) Index	(2) Employed	(3) Hours[c]	(4) Earnings[c]	(5) Hourly wage[c]	(6) Written contract
Public	0.120	0.052	0.201	0.338	0.197	0.020
	(0.027)	(0.012)	(0.052)	(0.074)	(0.040)	(0.010)
Placebo	0.027	0.020	0.039	0.069	0.054	0.005
	(0.043)	(0.027)	(0.075)	(0.184)	(0.129)	(0.021)
p: public = placebo	0.041	0.245	0.045	0.147	0.266	0.471
Placebo / public ratio	0.222	0.379	0.196	0.204	0.275	0.238
# observations	6609	6607	6598	6589	6574	6575
# clusters	84	84	84	84	84	84

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. Mean outcome is for the control group. All outcomes use a 7-day recall/forecast period unless marked with [a] (30-day recall/forecast period) or [b] (since treatment). Outcomes marked with [c] use the inverse hyperbolic sine transformation. The index in the first column shows the inverse covariance-weighted averages of the 5 labor market outcomes, following Anderson (2008). The mean ratio of placebo to public effects is 0.258 for all 5 non-index outcomes. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

F Experiments with Firms: Willingness to Pay and Skill Ranking

This appendix provides more information about the firm-facing experiments described in Sections 5.1 and 5.2. We recruit a sample of 69 firms located in commercial areas near the low-income residential areas in Johannesburg where most workseekers in our sample live. We survey them about their hiring practices, measure their willingness-to-pay for a database containing information about assessment results for workseekers in our sample, and measure their preferences for different types of skills using an incentivized resume-ranking exercise.

Table F.1 reports summary statistics for this sample. Table F.2 shows summary statistics on firms' preferences for different types of skills. Figures F.1 and F.2 shows screenshots of the platform marketed to firms. Figure F.3 shows the distribution of willingness-to-pay.

Table F.1: Summary Statistics for Firm Sample

Variable	# obs	Mean	Std dev.	10^{th} pctile	90^{th} pctile
Wholesale & retail trade	69	0.623	0.488		
Transport, storage & communication	69	0.014	0.120		
Restaurant & hospitality	69	0.188	0.394		
Agriculture	69	0.014	0.120		
Financial & insurance	69	0.087	0.284		
Community & social services	69	0.014	0.120		
Hiring decisions made exclusively at location interviewed	69	0.754	0.434		
# employees	69	15.0	29.6	3.0	32.0
# entry-level employees	67	7.24	14.94	0.00	14.00
# vacancies for entry-level employees	59	1.42	3.70	0.00	4.00
# entry-level hires expected in next 12 months	58	3.95	5.43	0.00	10.00
# applications received for last entry-level vacancy posted	56	16.2	21.2	2.0	30.0
# weeks required to fill last entry-level vacancy posted	58	4.17	6.47	1.00	8.00
Uses external recruiting services	69	1.75	0.43	1.00	2.00
Total payroll costs in last financial year	31	1,277,847	2,766,868	78,000	3,200,000
Mean monthly compensation for employees in last financial year	58	8,447	16,273	2,500	9,000

Table shows summary statistics for selected firm attributes variables. Percentiles are omitted for binary variables. First six rows are indicators for sectors. # observations varies due to item non-response. All monetary figures are reported in South Africa Rands. 1 Rand ≈ USD0.16 in purchasing power parity terms. Missing values for the final two variables are more common because the survey was completed by the person responsible for hiring decisions, who did not always have access to financial records.

Table F.2: Firm Ranking of Profiles with Different Assessment Results and Education

		(1)	(2)	(3)
Profile content		Share of firms ranking profile		Median
Top tercile	Highest education	First	Last	ranking
Communication	Complete secondary school	0.119	0.015	3
Concept formation	Complete secondary school	0.075	0.030	4
Focus	Complete secondary school	0.328	0.060	3
Grit	Complete secondary school	0.134	0.045	4
Numeracy	Complete secondary school	0.060	0.090	2
Planning	Complete secondary school	0.194	0.000	4
None	One-year post-secondary diploma	0.090	0.761	7

Table shows summary statistics from firms' ranking of profiles with different skill profiles and different level of education. All profiles have middle terciles for skills except that listed in the first column.

Figure F.1: Screenshots of Login Page and Filtering Page

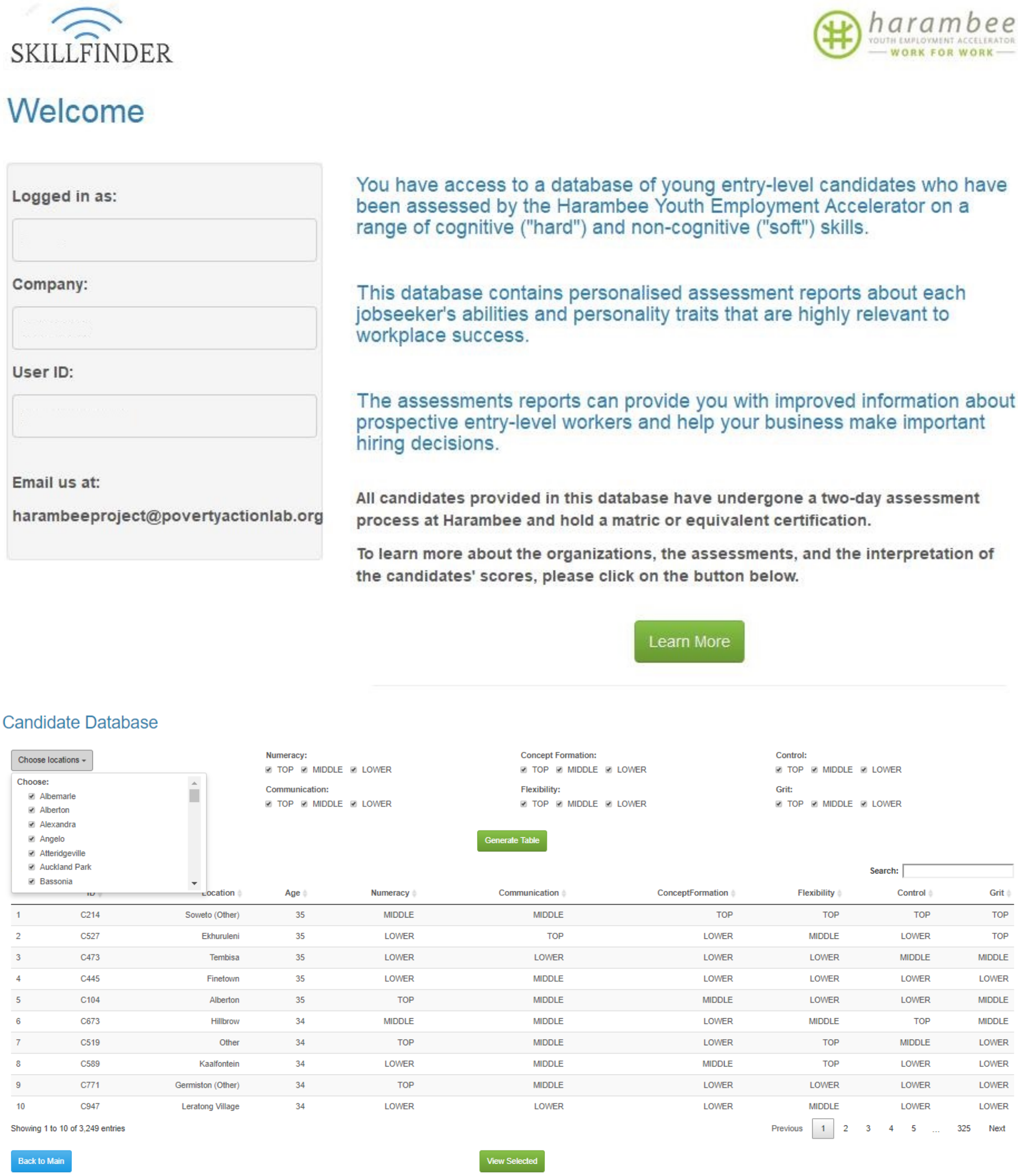
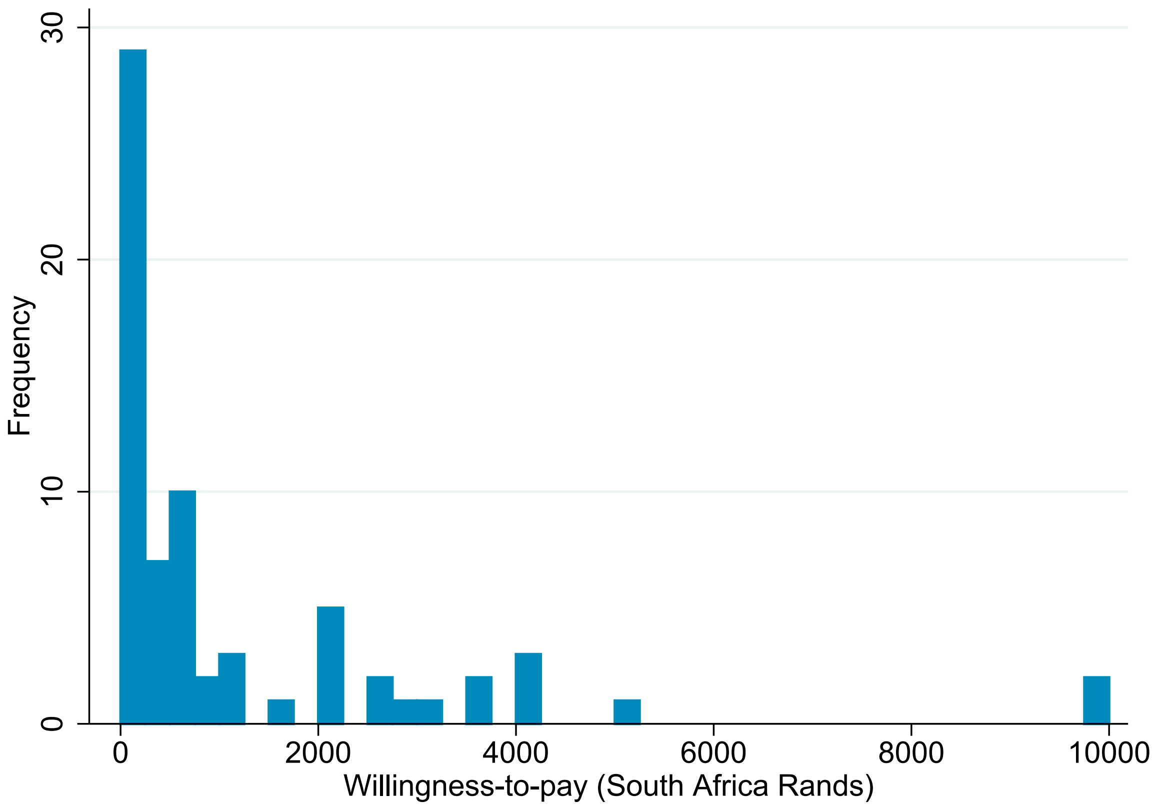


Figure F.2: Screenshot of Individual Candidate Profile on Platform

Figure F.3: Willingness-to-pay for Database of Workseekers' Assessment Results



Notes: This figure shows the distribution of willingness-to-pay for access to the database of assessment results described in section 5.1 and shown in Figures F.1 and F.2. Values are in South African Rands, with 1 Rand ≈ 0.16 US$ in purchasing power parity terms. The maximum possible bid is 10,000 South African Rands.