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

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

November 14, 2022

Abstract

We assess South African workseekers' skills and disseminate the assessment results to explore how limited information affects firm and workseeker behavior. Giving workseekers assessment results that they can credibly share with firms increases workseekers' employment and earnings and better aligns their skills, beliefs and search strategies. Giving workseekers assessment results that they cannot easily share with firms has similar effects on beliefs and search, but smaller effects on employment and earnings. Giving assessment results only to firms shifts interview decisions. These findings show that getting credible skill information to the right agents can improve outcomes in the labor market.

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

*We are grateful for helpful comments from Martin Abel, Peter Arcidiacono, Vittorio Bassi, Paolo Belli, Emily Breza, Rulof Burger, Stefano Caria, Jishnu Das, Giacomo di Giorgi, Taryn Dinkelman, Rebecca Dizon-Ross, Stefan Dercon, Erica Field, Gary Fields, Simon Franklin, Doug Gollin, Marek Hanusch, Rachel Heath, Lukas Hensel, Clement Imbert, Lisa Kahn, Julien Labonne, Clare Leaver, David McKenzie, Melanie Morten, Simon Quinn, Imran Rasul, Duncan Thomas, Gabriel Ulyssea, Ian Walker, Chris Woodruff, and seminar participants at the Barcelona Summer Forum, Berkeley, BREAD Fall Meetings, Duke, Georgia, Harvard/MIT, NBER Summer Institute, Notre Dame, NYU, Oxford, SoLE conference, 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, Brynde Kreft, Stephanie Palla, Rachel Sayers, Gabriel Cunha, and Jenn Kades 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 Gender Innovation Lab, and received funding from the National Science Foundation (# 1824413), Private Enterprise Development in Low Income Countries (# 3024 and # 4728), W. E. Upjohn Institute for Employment Research, and the UKRI GCRF Accelerating Achievement for Africa's Adolescents Hub. This study has been approved by the University of Cape Town Commerce Faculty Ethics in Research Committee, the Duke Institutional Review Board for the Protection of Human Subjects in Non-Medical Research (# D0368) and the Oxford Department of Economics Research Ethics Committee (# 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

Workseekers make job search decisions and firms make hiring decisions using potentially limited information about workseekers' skills. Limited information for firms can lead to hiring poorly-matched workers and to wedges between wage offers and productivity (Altonji and Pierret, 2001; Arcidiacono et al., 2010; Farber and Gibbons, 1996; Kahn and Lange, 2014). These hiring distortions can reduce both employment and average wages conditional on employment (Aigner and Cain, 1977; Pallais, 2014). Limited information for workseekers can lead them to search for jobs that poorly match their skills or withdraw from search entirely (Belot et al., 2018; Conlon et al., 2018). These search distortions can also lead to lower employment and lower wages conditional on employment. When both sides of the market receive credible information on workseekers' skills or past performance, these workseekers' labor market outcomes can improve (Abebe et al., 2020a; Abel et al., 2020; Bassi and Nansamba, 2020; Pallais, 2014). These information problems may be particularly important in settings where hiring is less formal and education provides less information about skills (Pritchett, 2013). Limited information may exacerbate other frictions in developing country labor markets, such as high search and migration costs (Abebe et al., 2020b; Bryan et al., 2014; Franklin, 2017).

We study how providing additional information about workseekers' skills affects job search, hiring, and workseekers' labor market outcomes. We run a series of field experiments that manipulate firms' and workseekers' information about workseekers' skills. We provide evidence that *both* firms *and* workseekers adjust their behavior when they acquire new information about workseekers' skills, suggesting they both face limited information. Their responses lead to substantial improvements in workseekers' outcomes in the labor market. The magnitude of effects suggests such information frictions may be an important target of government policy. The finding that both firms and workseekers respond to information is important both conceptually and for the design of information-provision products and policies. Many existing policies provide information directly to only one side of the market and may have different returns depending on whether this information can be shared with the other side of the market. For example, workseeker skill assessments offered in job search assistance programs can help workseekers to change job search strategies. But their returns may be different if the workseekers can also credibly share the assessment results with prospective employers. Most existing papers study only one side of the market or study simultaneous information revelation to both workseekers and firms.

We study firms' and workseekers' responses to learning workseekers' results on standardized skill assessments. The assessments measure non-specialist skills such as communication, numeracy, and grit and draw on existing tools used by job placement agencies and large firms. The 6,891 assessed workseekers are drawn from a population where limited information may be important. They are unemployed or underemployed black youths in urban South Africa with limited post-secondary education, work experience, and access to referral networks. This population faces statistical discrimination in this labor market and has limited information about labor market prospects (Banerjee and Sequeira, 2020; Malindi, 2017; Pugatch, 2018).

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

Second, we show that these labor market effects are smaller when workseekers cannot easily and credibly share assessment results with firms. To show this, we randomly select some assessed workseekers for a 'private' certification intervention. This intervention gives them one physical certificate that shows their assessment results and describes the assessments, but deliberately excludes features designed to make the public certificate credible to firms: branding and the workseekers' identifying information. Private and public certification have very similar effects on workseekers' beliefs about their skills and how they target job search based on their skills. But private certification has no effect on employment and raises earnings by less than public certification. The relative outcomes in the private certification and control groups suggest but do not prove that workseeker responses to additional information contribute to improved labor market outcomes. The relative outcomes in the public and private certification groups suggests an important role for firm responses to additional information and highlight the importance of getting credible information to firms as well as workseekers. A small share of workseekers also use private certificates in job applications, but this does not appear to explain the positive earnings effects of private certification.[1]

Third, we show that directly giving firms information about workseekers' skills changes their behavior. To show this, we run an audit/correspondence experiment that manipulates firms' information without 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

[1] The positive earnings effect of private certification is driven by workseekers who do *not* use the certificates in job applications. This suggests that the private effect on earnings is not primarily driven by information transmission to firms, consistent with the unbranded, unidentified design of the private certificates. However, we cannot rule out some role for information transmission to firms.

vacancy, randomizing whether applications include public certificates. When only one application sent to a vacancy includes a certificate, that application is 11% more likely to get an interview than the applications without certificates. But this benefit vanishes as the vacancy gets more applications with certificates. This pattern is consistent with firms acting on information from skill certification, although their actions may depend on the scale of certification.

These three experiments demonstrate our main finding: additional information about workseekers' skills improves labor market outcomes, but it matters who gets this information and how credibly and easily it can be shared. In addition, we present four secondary findings about the role of limited information in this labor market. These findings rely on heterogeneity analysis and smaller experiments and we interpret them as suggestive rather than conclusive. First, learning specific assessment results is important, not just learning that workseekers have been assessed. The certification effects are not driven, for example, 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, the certification effects are more consistent with horizontal than vertical differentiation of workseekers: certification helps firms identify which workseekers are suited for different jobs more than it helps firms identify a subset of workseekers suited for all jobs.[2] This may occur because, in this context, preferences for different skills vary across firms and relative performance in different assessments varies across workseekers. 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, although we do not directly observe if certified workseekers become employed at the expense of workseekers outside our sample, most of our results are consistent with economic mechanisms in which certification can increase total employment.

Our main contribution is to study workseeker and firm responses to additional information about workseekers' skills, highlighting the importance of information that is easily and credibly shared. This extends existing work documenting labor market patterns consistent with either firms alone or workseekers alone facing limited information, in both developed and developing economies.[3] We build on this work by showing that *both* firms and workseekers in the same labor market respond to new information about workseekers' skills. Our work is most similar to studies that provide

[2]This finding is consistent with recent work on information frictions in matching models with multidimensional skills (Fredriksson et al., 2018; Guvenen et al., 2020; Lise and Postel-Vinay, 2020).

[3]Altonji and Pierret (2001), Arcidiacono et al. (2010), Farber and Gibbons (1996), and Kahn and Lange (2014) show that wages align more closely with skills as tenure increases, consistent with firms' facing limited information about skills at the time of hiring. Wage and retention patterns for workers hired through referrals are also consistent with limited information (Ioannides and Loury, 2004; Heath, 2018) and some researchers find that workers have better labor market outcomes when they have formal educational qualifications, conditional on measured skills (Alfonsi et al., 2017; MacLeod et al., 2017). Workseekers can have systematically inaccurate beliefs about their labor market prospects (Spinnewijn, 2015) and their job search decisions can change when they learn more about their prospects (Ahn et al., 2019; Altmann et al., 2018; Banerjee and Sequeira, 2020; Belot et al., 2018), although these papers do not specifically examine limited information about workseekers' skills.

information to both firms and workseekers about skill assessment results (Abebe et al., 2020a; Bassi and Nansamba, 2020; Groh et al., 2015) or evaluations from workseekers' past employers (Abel et al., 2020; Pallais, 2014). We build on this work by experimentally varying which agents receive the information and how credibly and easily it can be shared.[4]

Understanding how different agents respond to information is important for designing mechanisms that private actors or governments can use to address limited information. Separate firm-facing and workseeker-facing mechanisms are common, but their effects may depend on the information available to the other side of the market. For example, on the workseeker side, some job search assistance programs offer skill assessments to workseekers (McCall et al., 2016). This can inform workseekers and improve their search targeting. But if firms do not learn these assessment results, then firms' hiring choices and wage offers will remain distorted and workseekers' improved search will have limited returns. On the firm side, skill assessments are sometimes used to inform firm hiring decisions (Autor and Scarborough, 2008; Hoffman et al., 2018). But if workseekers have limited information, they might not apply for jobs that match their skills, leaving firms to assess and select from a sub-optimal pool of applicants. Alonso (2018) shows theoretically that giving better information to only firms or only workseekers in labor market matching can reduce welfare when they cannot or will not share that information with the other side of the market.

Second, this paper complements work on the aggregate implications of limited information in the labor market. Canonical models show that search and matching frictions facing individual workseekers and firms can generate aggregate unemployment (Mortensen and Pissarides, 1999). Our findings offer an experimental foundation for general equilibrium models that show how either firms' or workseekers' limited information about match productivity can distort aggregate employment (Jovanovic, 1979; Gonzalez and Shi, 2010). In particular, our findings complement work by Donovan et al. (2018), who show that models with limited information about workseekers' skills can explain aggregate labor market dynamics in developing countries. We borrow the language of the search and matching literature, referring to distortions in workseeker and/or firm behavior due to limited information as 'information frictions.'

Third, our findings on information frictions are relevant to the design of active labor market programs (ALMPs). We show that a skill assessment and certification intervention, delivered during recruitment for an ALMP, can substantially and cheaply improve participants' employment and earnings. The employment effect is almost three times larger than the mean effect size of the active labor market programs reviewed by Card et al. (2018). The average earnings gain in the first three months after treatment is 5.6 times the average variable cost of adding this certification intervention onto an existing assessment program and 2.3 times the average variable cost of assessment and

[4] Abel et al. (2020) reveal reference letters to both sides of the market and only to firms, but do not measure workseeker belief updating or search targeting. Related work by Banerjee and Chiplunkar (2020) studies the implications of university placement officers having limited information about workseekers' preferences over job types.

certification. [5] Skill assessment and certification may enhance the value of ALMPs to participating workseekers even when other mechanisms for learning about workseekers' skills exist. Importantly, certification is available to first-time workseekers, unlike reference letters or performance evaluations from past employers (Abel et al., 2020; Pallais, 2014). Assessment results can be certified to multiple employers, while workplace performance at one employer may be imperfectly observed by other employers (Kahn, 2013). Certification can help workseekers excluded from referral networks or firms who receive referrals based on factors poorly aligned with workseekers' skills (Beaman and Magruder, 2012; Beaman et al., 2018; Chandrasekhar et al., 2020).

We describe the economic environment in Section 2: a simple conceptual framework, the context, the 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 limited information. In Section 5, we discuss secondary results about what workseekers and firms learn from certification, what this implies for the effects of certification on different types of workseekers, and what this might imply for certification at a larger scale. We conclude in Section 6 and discuss questions around markets for assessment-based certification.

2 Economic Environment

2.1 Conceptual Framework

In this section, we sketch a simple conceptual framework with two goals. First, the framework illustrates how either workseeker- or firm-side limited information can lower two labor market outcomes: the employment rate and the mean wage conditional on employment. Hence, observing that employment and/or wages rise when both firms and workseekers have access to more information does not show which side(s) of the market responds to information. Second, the framework illustrates the mechanisms that link limited information to distortions in firm and workseeker behavior and hence to lower wages and employment. This guides our empirical tests of these mechanisms.

Consider a stylized economy consisting of infinitely many type W_1 and W_2 workseekers and type J_1 and J_2 jobs. Workseekers may choose not to search, to search for type 1 jobs, or to 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 a job with probability $P_{i,j}$. Conditional on meeting, the match produces output with pecuniary value $V_{i,j}$ net of any screening cost the firm incurs during hiring and pays wage $W_{i,j} \leq V_{i,j}$. The workseeker receives utility $P_{i,j} \cdot U(W_{i,j}) - C$ if she searches and zero otherwise, implying that she has a reservation wage $\underline{W}_i(C, P)$.[6] There will be some labor

[5] In a similar spirit, several papers show that making low-cost changes to ALMPs so they provide more information to firms and/or workseekers improves their effectiveness (Abel et al., 2020; Belot et al., 2018; Wheeler et al., 2019).

[6] 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.

force non-participation if search costs are high relative to the expected utility of working and some unemployment if the meeting probability $P_{i,j}$ is less than one for some (i, j).

We make some additional simplifying assumptions for this discussion, but none of the results in the framework depend on these additional assumptions. First, we assume fraction p of all workseekers and all jobs are type 1. Second, we assume that 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, and similarly for type j workseekers and type j jobs. Under these assumptions, if type i workseekers choose to search, they will choose to search for type i jobs rather than type j jobs, and vice versa.

Either firms or workseekers can have limited information about workseekers' skills in this environment. First, we consider the case where only firms observe workseekers' types with error. This can occur if attributes observable to firms, like educational qualifications or past work experience, provide limited information about skills. Workseekers search for the 'right' types of jobs but firms do 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 or severance pay and dispute resolution costs when firing workseekers. This reduces mean wages conditional on employment and, if offered wages for some vacancies are below the reservation wage or a legal minimum wage, reduces the employment rate. If firms have access to screening technology, they may observe workseekers' types more accurately and be able to pay workers a larger share of the match output. But the cost of the screening technology presumably reduces net match output, so the value available to pay in wages remains lower than in a world with perfect information. Aigner and Cain (1977) and Pallais (2014) prove results of this flavor formally.

Second, we consider the case where only workseekers observe their types with error. This can occur if, for example, workseekers receive limited information about their own type from education or work experience or if they have little education or work experience. In this case, each workseeker chooses whether and where to search based on her perceived type. If a type i workseeker 'incorrectly' searches for a type $j \neq i$ job, she is less likely to meet a firm and, conditional on meeting a firm, will produce less and earn a lower wage. This reduces mean wages conditional on employment by generating some mismatches between workseeker and job types. 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 mismatches between workseeker and job types may not generate enough output to offer wages above the reservation wage or minimum wage. The former mechanism can occur if, for example, firms offering different types of jobs hire using different channels, like posting formal adverts versus hiring walk-ins. The latter mechanism can occur if, for example, search costs and hence reservation wages are high or there is a legal minimum wage. Belot et al. (2018) and Falk et al. (2006) prove results of this flavor formally.

This simple framework shows that observing a rise in employment and/or wages when both firms and workseekers acquire more information does not show which side(s) of the market faces limited information. This highlights the importance of studying both firm and workseeker responses to new information. Depending on the structure of the model, limited information on both sides of the market might interact to generate larger distortions or partly offset each other.[7] We focus on the static case for simplicity, but recognize that the effect of limited information may differ in a dynamic framework with learning by firms or workseekers (Conlon et al., 2018; Lange, 2007).

The framework allows either horizontal or vertical differentiation of workseekers. 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 limited information can lower the employment rate and the mean wage conditional on employment. With horizontal differentiation, limited information on either side of the market can lower wages conditional on employment for all types of workseekers. With vertical differentiation, firm-side limited information can lower wages for type i workseekers mistaken for type j workseekers and raise 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, as there are few sources of information on workseekers' skills. Many young workseekers have no work experience several years after completing education, limiting the scope to learn about or signal their skills through experience (Ingle and Mlatsheni, 2017). 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). Workseekers who have completed secondary school typically report their grades in the nationally standardized graduation exam 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).[8] This limits the scope for firms and workseekers to learn about workseekers' skills from their educational attainment. Certification

[7]We do not explore interaction effects in detail in this paper. They are not identified by our experimental design without strong assumptions, because we do not directly cross-randomize the information available to firms and workseekers. They are also difficult to characterize theoretically because they depend on second-order beliefs, which we do not observe. For example, firms' return to investing in screening technology depends on their own uncertainty about workseekers' skills and their beliefs about the workseekers' own uncertainty about their own skills and what this implies for their search decisions.

[8]The limited information content of education qualifications is consistent with the large role of referrals in hiring: more than half of all firms report that referrals are their preferred recruitment mechanism (Schöer et al., 2014).

can thus provide both firms and workseekers with additional information on workseekers' skills.

Second, 'wrong' hires are costly to firms. Firing a worker requires a complex and lengthy process and can be challenged by even temporary employees in courts and specialized dispute resolution bodies.[9] 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.[10] Consistent with these factors, giving firms free consulting on labor regulation increases hiring (Bertrand and Crépon, 2019).

Third, reservation and legal minimum wages exist. Minimum wage compliance in the formal sector is high (Bhorat et al., 2016; International Labour Organization, 2016). Commuting costs are high and likely to raise reservation wages (Kerr, 2017). The nearly universal state pension system gives workseekers in multi-generation households access to non-labor market income, increasing reservation wages (Abel, 2019).

Fourth, employment rates are low. In our study period, unemployment in Johannesburg was 28% for the working-age population, 51% for ages 15-24, and 32% for ages 25-34 (Statistics South Africa, 2016b).[11] Low employment in the presence of information frictions, costs from 'wrong' hires, and reservation or minimum wages are consistent with our conceptual framework. Particularly low employment for youths is also consistent with information frictions, as youths have less of the search and work experience that could 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, active workseekers from low-income backgrounds with limited work experience. 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. This is a sample from a policy- and theory-relevant population likely to face information frictions, rather than a population-representative sample.

To recruit the sample, we work with the Harambee Youth Employment Accelerator, a social enterprise that assesses the skills of inexperienced workseekers and matches them to employers looking for entry-level workseekers, among other activities aimed at addressing a mismatch of demand and supply in the South African youth labor market. Harambee recruits candidates through

[9] Small firms 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).

[10] Only 18% of SME owners know the conditions that made a contract valid or rules governing severance pay (Bertrand and Crépon, 2019).

[11] 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 in full-time education or not in the labor force.

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.[12] Eligible candidates are invited to two days of standardized skill assessments. Some candidates are invited to further job readiness training based on their assessment results and residential location, but only 0.2% of candidates in our sample get jobs through this training. Our sample consists of all candidates who arrive at Harambee for the second of these two testing days, on 84 operational days.

We conduct three surveys to measure workseekers' labor market outcomes, job search, and beliefs about their skills and the labor market. The baseline is a self-administered questionnaire that candidates complete on desktop computers at Harambee under supervision. This is administered after candidates have done skills assessments but before they receive information about their results. We collect endline data in a 25-minute phone survey 3-4 months after treatment.[13] The phone survey response rate is 96%, leaving an endline sample of 6,609 respondents. The response rate is balanced across treatment groups (Table D.6) and unrelated to most baseline covariates (Table D.7). We also conduct 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 the 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 D.1 and D.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. Malindi (2017) shows that young, black workseekers with relatively low levels of education face discrimination in this labor market, with wage dynamics consistent with information frictions and statistical discrimination.

Of the sample, 38% worked in the week before the baseline and 70% had ever worked, but only 9% had ever held a long-term job. Conditional on working, mean weekly earnings in the week before the baseline was 565 South African rand (94 USD in purchasing power parity terms), slightly below the minimum wage for a full-time worker in most sectors. 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.

Of the sample, 97% searched for work in the week before the baseline. In that week, they spent on average 17 hours and 242 South African rand (40 USD PPP) searching. The relatively high search costs suggest that welfare gains for workseekers are possible from improved search targeting.

[12] 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 12 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.

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

Workseekers submitted on average 10 applications in the preceding month and received 1.2 offers. 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.

Unsurprisingly, this sample is positively selected on search behavior and negatively selected on labor market outcomes. We establish this by comparing our sample to people from the same city with the same distribution of age, education, gender, and race using South Africa's nationally representative Quarterly Labour Force Survey (Statistics South Africa, 2016a, 2017). Our sample has roughly the same employment rate but earns only 25% as much, potentially reflecting both lower hours and lower hourly wages, and is roughly twice as likely to be searching for work.

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. Client firms have paid Harambee to screen roughly 160,000 prospective workers using these assessments. Appendix A describes each assessment in detail, their psychometric properties, and how some Harambee client firms use them in hiring.

Each assessment session is led by two or three industrial psychologists, who manage a team of facilitators. Assessments are conducted in English and are self-administered on desktop computers. Appendix Table D.1 shows standardized scores on the assessments. There is a fairly even spread of candidates over the distribution and little evidence of ceiling effects.

Appendix Table A.2 shows the correlation matrix between different skills. 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 horizontally differentiate candidates based on their relative skills rather than only ranking or vertically differentiating them in a single dimension of skills.

Candidates have inaccurate beliefs about their own types, suggesting a role for workseeker-side information frictions. 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 29% 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 for all three assessments (Appendix Table D.1).

Workseeker 'types' in our data are multidimensional and ordinal within each dimension, rather than the simple case of binary types discussed in the conceptual framework. This means that workseekers may have inaccurate beliefs because they imperfectly observe the population distribution of skills, even if they perfectly observe their own skills.

3 Labor Market Effects of Skill Certification

3.1 Intervention

Our first certification intervention gives candidates information about their assessment results that they can easily and credibly share with prospective employers. The effects of this intervention may reflect changes in firm- or workseeker-side behavior. In either case, the framework predicts that certified workseekers will have higher employment and higher earnings conditional on employment.

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. To provide credibility to the assessments and results, the certificate is branded with the World Bank and Harambee logos. Harambee is a widely recognized brand in South African marketing surveys (Mackay, 2014).

The certificate describes the skills measured by each assessment. The certificate directs the reader to 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. For each skill, the certificate shows the tercile in which the candidate ranked on each assessment, compared to other candidates assessed by Harambee.[14] The candidates assessed by Harambee are described as South African high school graduates aged 18-34 from disadvantaged backgrounds. To link candidates with certificates, each certificate shows the candidate's name and national identity number. National identity numbers are typically shown on resumes and school transcripts in South Africa.

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. To ensure briefings were standardized, 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.

3.2 Experimental Design

We randomly divide our workseeker sample into a public certification group, a control group, and other groups discussed in the next section of the paper. We randomize treatment by assessment date to reduce risks of spillovers between treated and control workseekers, assigning 2,247 workseekers

[14] In piloting, both workseekers and firms found certificates with only rankings easier to interpret than certificates with only cardinal scores or both rankings and 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:

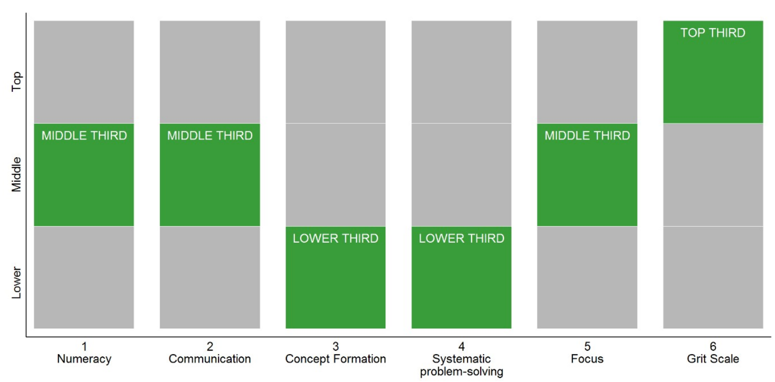
1. The Numeracy tests measure candidates' ability to apply numerical concepts at a National Qualifications Framework (NQF) level, such as working with fractions, ratios, money, percentages and units, and performing calculations with time and area. This score is an average of two numeracy tests the candidate 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.

This figure shows an example of the certificates given to candidates in the certification treatment. Each certificate shows some information about the assessments, the candidate's 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 copy of the certificate, and guidelines on how to request more certificates.

assessed over 27 days to certification and 2,274 workseekers assessed over a different 27 days to the control group. Treated and control workseekers differ in only one way: treated workseekers receive the certification intervention described above, while control workseekers receive no information about their assessment results and no certificate to enable them to share results with firms. Control workseekers received the same experience at Harambee as all workseekers before the experiment and were not told that workseekers assessed on other days received certificates, so it is unlikely that workseekers assigned to the control group were discouraged or inferred anything about the assessment results from not getting certificates. 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.[15] This differs from the design in Abebe et al. (2020a), where treated workseekers receive both 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, and $\mathbf{X}_{id}$ is a vector of prespecified baseline covariates. S_d is a block fixed effect, to account for the fact that we randomly assign days to treatment groups within blocks of 6-10 sequential days. 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 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 labor market outcomes, which allows us to examine on the distribution of outcomes for employed candidates.

The estimating equations and variable definitions are prespecified. All outcomes whose treatment effects are reported in tables/figures are prespecified except where we indicate otherwise, although not all outcomes discussed in the text are prespecified. Our estimates of key treatment effects are robust to omitting the prespecified covariates (Table D.8) and to including the covariates that are unbalanced at baseline (Table D.9). Our inferences about key treatment effects are robust to

[15] 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 survey, 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	0.337	0.197	0.020
	(0.012)	(0.052)	(0.074)	(0.039)	(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 outcomes are for the control group. All outcomes use a 7-day recall period. Outcomes marked with [c] use the inverse hyperbolic sine transformation for the treatment effects but the control group means are reported in levels. All monetary figures are reported in South Africa Rands. 1 Rand ≈ USD 0.167 in purchasing power parity terms. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

several adjustments for multiple testing: estimating q-values that control the false discovery rate across related outcomes, combining related outcomes into indices, and estimating q-values across indices following Anderson (2008) and Benjamini et al. (2006).

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 30.9 percentage points (Table 1, column 1). We also ask about employment in each calendar month between treatment and endline and show in Table D.12 that certification increases employment in every month between treatment and follow-up.

Certification increases average weekly hours worked, coded as zero for non-employed candidates, by 20% (column 2). The treatment effect on hours may reflect two effects: an extensive margin effect if treatment increases the employment rate and an intensive margin effect if 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 C). 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 average hours as the inframarginally employed untreated candidates. We define the intensive margin effect as the difference between the total treatment effect on hours and the extensive margin effect on hours. Intuitively, this is the treatment effect on hours due to changes in hours worked conditional on employment. We find that the entire effect on hours is explained by the extensive margin effect (Table 2, column 1). This shows that treated candidates do not work longer hours conditional on employment, but are simply more likely to be employed.

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	0.337	0.197	0.020
	(0.052)	(0.073)	(0.039)	(0.010)
Extensive margin	0.188	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)	(0.027)	(0.008)
Treatment effect conditional	0.037	0.194	0.158	-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 margin effects. The extensive margin effects 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 margin effects are the differences between the treatment effects and extensive margin effects, which must be due to changes in job characteristics for the employed candidates 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. Outcomes marked with ^{c} use the inverse hyperbolic sine transformation.

The second main effect of certification is to increase earnings. Weekly earnings increase by 34% (Table 1, column 3). The extensive margin effect and intensive margin effects account for respectively 27 and 7 percentage points of the 34% increase in earnings (Table 2, column 2). Hourly wages, calculated by dividing earnings by hours, also increase by 20% (Table 1, column 4). The extensive and intensive margin effects account for respectively 14 and 6 percentage points of the 20% increase in wages (Table 2, column 3). These results show that treatment increases earnings mainly by increasing employment, but also increases earnings conditional on employment.

These results are consistent with the conceptual framework: more information about workseeker skills (i.e., types) increases the latent value net of screening costs of some workseeker-job matches, leading to higher employment and mean earnings conditional on employment. However, these results do not pin down the relative contributions of lower screening costs and higher match quality without either data on firm screening activities or stronger assumptions. Treatment also increases another common proxy for match quality: it increases average tenure at endline by 0.1 months (standard error 0.04 months). However, the the 3-4 month period between baseline and endline is too short to infer a strong relationship between match quality and tenure.

These results allow us to reject a special case of the framework where more information increases only job-finding rates but not the value of firm-worker matches net of screening costs. In this special case, treatment would not increase earnings conditional on employment. This special case does not match the positive treatment effects we find on earnings and wages conditional on employment, nor does it match the quantile treatment effects on earnings shown in Figure 2. In this special case, the quantile treatment effects would be large and positive from the 66th to 71st percentiles where the marginally employed workseekers went from zero to positive but low earnings, and zero for all other percentiles. Instead, we see positive quantile treatment effects from the 66$^{\text{th}}$ percentile upward, although they are not statistically significantly different to zero above the 93$^{\text{rd}}$ percentile.

Finally, certification shifts the types of employment. Certification increases the probability of having a written contract, Statistics South Africa's definition of a formal job, by 2 percentage points (Table 1, column 5). This effect is entirely explained by the higher employment rate (Table 2, column 4). Furthermore, 4 percentage points of the 5.2 percentage point increase in employment are in wage employment, and only 1.2 percentage points are in self employment. The wage employment and formality results show that certification is particularly effective at getting workers into more formal jobs, which are more likely to use formal hiring processes where certificates can play a role.

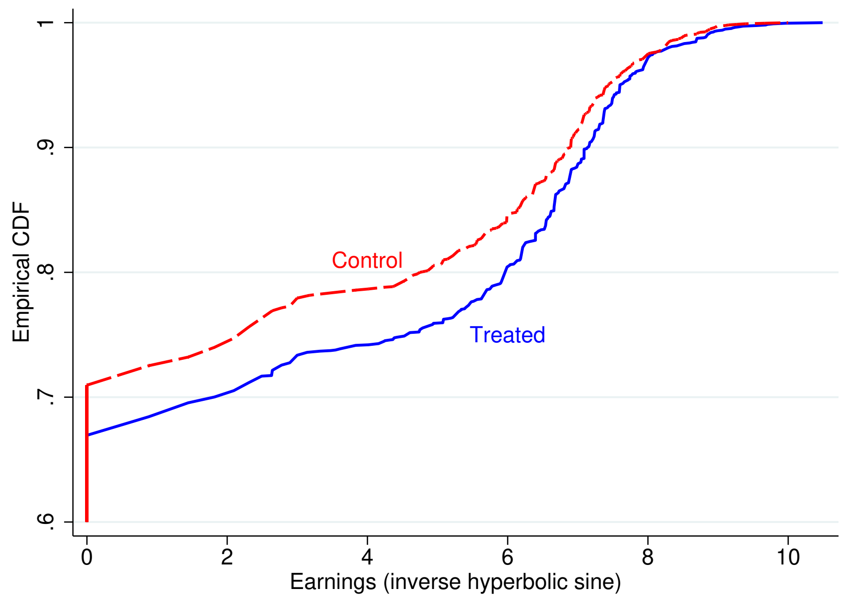
The effects on employment and earnings are substantial and easily exceed the cost of the program. The employment effect is almost three times larger than the mean standardized short-run effect size of active labor market programs reviewed by Card et al. (2018), larger than the effect of an intervention that helped similar South African workseekers get reference letters from past employers (Abel et al., 2020), and similar to the effect of a program that subsidized firms to hire South African workseekers from similar backgrounds (Levinsohn et al., 2013). The average earnings gain in the first three months after treatment is 778 South African rand (USD 130 PPP) – 5.6 times the average variable cost of adding certification onto an existing assessment program alone and 2.3 times the average variable cost of assessment and certification (details in Appendix B). The average weekly earnings gain is equal to 17% of the weekly adult poverty line in South Africa (details in Appendix D.2).

4 How Do Different Agents Respond to Certification?

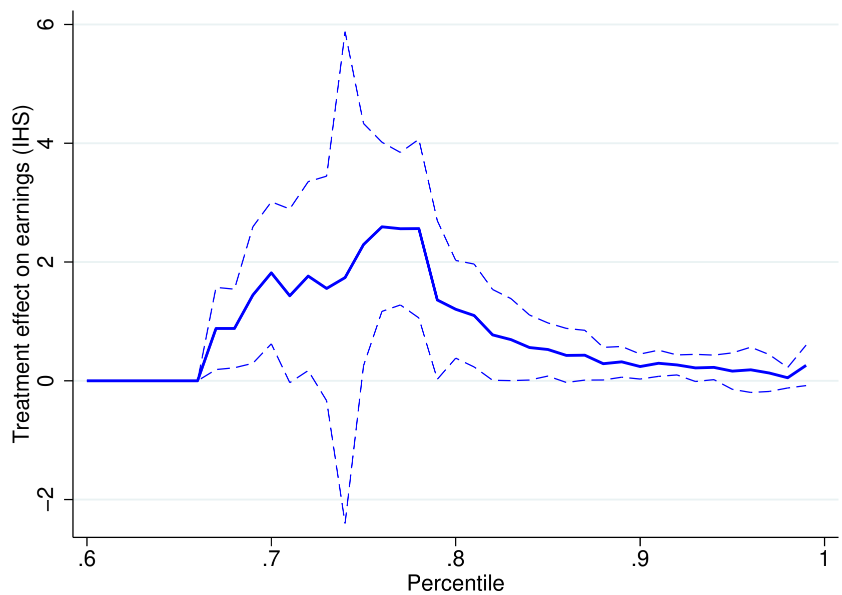
Certification may increase employment and earnings by providing information to firms, to workseekers, or to both sides of the market. This distinction matters for modeling information frictions and designing government or market-based remedies to limited information. In this section, we show that both sides of the market change behavior in response to new information and explore how these changes relate to the labor market effects of additional information. Our argument proceeds in three steps. First, we show that public certification changes workseekers' beliefs and search behavior in multiple ways. These changes don't conclusively show whether firms, workseekers, or both face limited information, motivating the second and third steps of the argument. Second, we discuss another arm of our workseeker experiment that reveals information to workseekers without helping them share the information with firms. The results of the different experimental arms show that both firms and workseekers face limited information and suggest but do not prove that both firms' and workseekers' behavioral changes upon receiving new information contribute to improved labor market outcomes. Third, we discuss an audit-style experiment that reveals information only to firms. The results of this experiment are consistent with firm-side limited information.

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



Panel A shows the empirical distributions of earnings in the control and public certification groups. Earnings are the inverse hyperbolic sine transformation of earnings in South African rand, with 1 rand $\approx$ 0.167 USD 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^{th} percentile because earnings below that value are zero. 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.

4.1 Public Certification Changes Job Search and Beliefs

We document three patterns in the effects of certification on workseekers' beliefs and job search behavior and then interpret these patterns. First, certification shifts workseekers' beliefs about their skills closer to their measured skills. We ask candidates if they think they scored in the top, middle, or bottom third on each of the six assessments, compared to other candidates assessed by Harambee. Certification increases the fraction of assessments where candidates' self-assessments match their measured results from 0.39 to 0.55 (Table 3, column 1). [16] In contrast, certification has no effect on candidates' self-esteem (column 2). This shows that their updated beliefs about the skills do not lead to more general updating about their self-worth.[17]

Second, certification 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).[18]

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 with certificates per candidate.[19] Applications with certificates generate an average of 0.43 interviews and 0.11 job offers over the 3-4 months from treatment to endline.

We interpret these patterns as evidence for limited information on both sides of the market. The first two patterns suggest a role for workseeker-side information frictions: candidates align their beliefs and job search more closely to their assessment results, potentially leading to better outcomes in the labor market. The third pattern suggests a role for firm-side information frictions: candidates use certificates with job applications, potentially making the applications more informative to employers, leading to more job interviews and offers. Jointly, these patterns lead candidates to expect 11% more offers in the next month, from a control group mean of 4.2 offers (column 8), and generate the improved outcomes in the labor market discussed in Section 3.3.

Before proceeding to the next experiments, we note that certification does not change multiple

[16] This question measures candidates' beliefs about their results in each assessment. These may differ from candidates' beliefs about their skills, if they believe the assessments are poor measures of their skills. Reassuringly, we obtain similar results when we measure candidates' beliefs about their skills in numeracy, communication, etc. rather than their results in these specific assessments. See Appendix D for details.

[17] Certification also has no effect on the distribution of self-esteem (Figure D.1) and has zero effects even for candidates who learn that they performed substantially worse or better on assessments than their baseline beliefs.

[18] We ask candidates separately about the skill demand of the jobs they target and about their perceived skills in two different parts of the survey. We construct the measure of search targeting from these two questions. This may be less susceptible to experimenter demand effects than asking them directly if their job search aligns with their skills. The result is similar for the fraction of candidates searching for jobs that most value the assessment in which they *think* they scored highest. This search targeting measure is not prespecified.

[19] The 6.7 additional applications with certificates follows from the 1.682 unit effect on the inverse hyperbolic since of the number of applications in column 5, and the fact that control workseekers send zero applications with certificates.

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	0.002	0.051		
	(0.008)	(0.013)	(0.010)		
Private certification	0.123	-0.002	0.047		
	(0.008)	(0.015)	(0.010)		
p: public = private	0.000	0.812	0.701		
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	0.290	0.572	0.144	0.036	0.054
	(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) Employed	(10) Hours[c]	(11) Earnings[c]	(12) Hourly wage[c]	(13) Written contract
Public certification	0.052	0.201	0.337	0.197	0.020
	(0.012)	(0.052)	(0.074)	(0.039)	(0.010)
Private certification	0.011	0.066	0.162	0.094	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 outcomes are 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). All outcomes use a 7-day recall period. Outcomes marked with [c] use the inverse hyperbolic sine transformation for the treatment effects but the control group means are reported in levels. All monetary figures are reported in South Africa Rands. 1 Rand $\approx$ USD 0.167 in purchasing power parity terms. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

prespecified measures of job search effort in the month before the endline survey: the probability of doing any search, number of applications submitted, hours spent searching, and money spent on search (Table D.12).[20] 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 extensive or intensive margin search effort but the endline may occur too late to detect this change. Employment already rises in the first month after treatment (Table D.12). This suggests that any changes in workseeker behavior that increase employment occur soon after treatment. The search effort questions use 7- or 30-day recall periods, which miss the period soon after treatment when candidates may have increased effort and found jobs. The questions on certificate use ask about the entire period between treatment and the endline survey, which will capture any short-term changes in search behavior.[21]

4.2 Workseekers Respond to Information That Is Difficult to Credibly Share with Firms

In this section, we show that workseekers' beliefs and search behavior change when they get more information about their skills, even when this information cannot be easily and credibly shared with firms. The specific pattern of results suggests, but does not conclusively prove, that these workseeker-side changes contribute to improvements in labor market outcomes.

To show this, we implement a 'private' certification intervention, distinct from the 'public' certification intervention described above. 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). The private treatment is designed to primarily provide information to the workseekers about their own skills.

Candidates in this group receive only one black-and-white, unbranded certificate, printed on low-quality paper, and do not receive an electronic version. 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 suggest that this is possible, unlike the briefing for 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. We assign 2,114

[20] We also find no heterogeneity in the treatment effects of certification on employment by any measure of baseline search effort (applications, cost, time, or indices that combine these measures). This may reflect the relatively high baseline search activity in our sample, with 97% of the participants actively searching in the week before baseline.

[21] 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.

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.

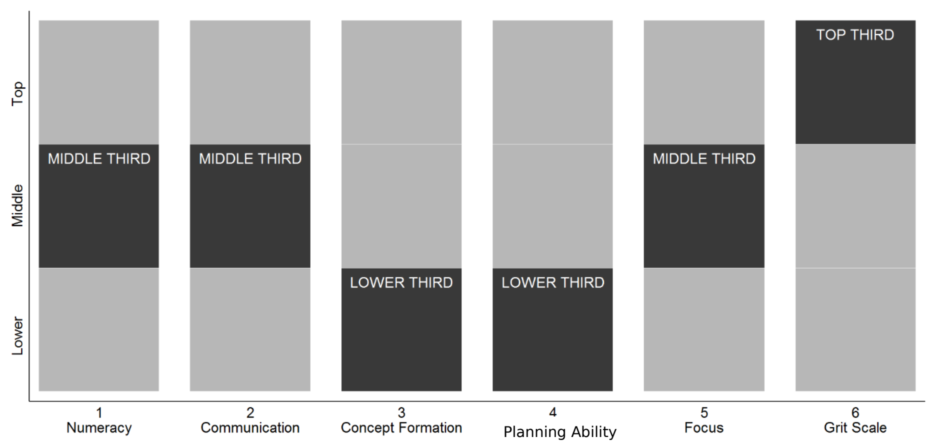
1. The Numeracy tests measure various maths abilities. Your score is the average of the two maths tests you did today at Harambee.
2. The Communication test measures English language ability through listening, reading and comprehension.
3. The Concept Formation test measures the ability to understand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

candidates assessed over 27 assessment days to private certification. We simultaneously randomize days to public certification, private certification, and control. The three groups are balanced on baseline characteristics (Table D.3).

The private and public certification interventions have similar effects on workseekers' beliefs and search targeting. Private certification makes workseekers' beliefs about their own skills 12 percentage points more accurate and has no effect on self-esteem (Table 3, columns 1-2).[22] Private certification increases search targeting by 5 percentage points, almost exactly the same magnitude as the public certification effect (column 3). Candidates in the private arm expect to receive 5% more offers than control candidates, significantly less than the 11% increase in expected offers in the public arm (column 8). This suggests that workseekers view the new information as useful, but less useful than when it is publicly certified and hence easy to credibly share with firms.

The private certification intervention has substantially 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 from zero, and statistically significantly smaller than the public certification effects (columns 9-10). Private certification increases earnings and hourly wages by respectively 16 and 9% but both effects are less than half the size of the public certification effects and are statistically significantly smaller (columns 11-12). The private certification effect on earnings is driven by workseekers who were not employed at baseline, so it reflects workseekers either getting or accepting higher-paying job offers, rather than using their new information to bargain up earnings at their current job. The intensive margin private certification effect on earnings of 0.103 inverse hyperbolic sine points is insignificantly larger than the equivalent public certification effect of 0.069 (Table D.14). This is not explained by differential selection between the two groups into employment on skill, education, work experience, or demographics.

We interpret the average treatment effects of the public and private certification interventions as evidence that both firms and workseekers face limited information and that providing more information leads to quantitatively important improvements in labor market outcomes. These improvements may reflect both firm- and workseeker-side learning, which we view as the most plausible interpretation. However, they may instead reflect only firm-side or only workseeker-side learning, possibilities we now discuss in turn.

First, we consider the possibility that the improvements in labor market outcomes from both

[22] The private certification effect on beliefs about skills is slightly smaller than the public certification effect. The former effect 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 keep copies of the public certificate 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 that workseekers' beliefs update in the same way straight after receiving the certificates and that the difference in the endline survey 3-4 months later is due to differential retention. See Appendix D and Table D.10 for details.

interventions are driven by only firm-side learning. This could occur if information from private certification 'leaks' to firms, explaining the earnings gap between the private certification and control groups. Some information does indeed leak to firms: 29% of workseekers used private certificates in job applications, sending an average of 1.8 applications with certificates and getting an average of 0.04 job offers from these applications (Table 3, columns 4-7). However, information transmission to firms is limited: workseekers use public certificates in job applications four times more than private certificates and get offers from applications with public certificates three times more often than from applications with private certificates.[23] Workseekers are also ten times more likely to include public certificates than private certificates with the applications they send us to participate in the audit study described in the next section. Even if information leaks to firms, it may have low credibility: private certificates do not have the candidate's name and identity number, so they cannot be linked to a particular candidate; have no branding from Harambee or the World Bank; do not explain that Harambee has used these assessments widely to place candidates with companies or that assessments predict workplace productivity; and do not link to a website. None of the 15 hiring managers interviewed during piloting reported that they would view the private certificates as credible. Furthermore, the positive earnings effect of private certification is driven entirely by workseekers who do *not* use the certificate with job applications. We interpret this result with caution, because it involves stratifying by a post-treatment outcome. But it suggests that some other mechanism helps to explain the earnings gap between the private certification and control groups, such as workseekers learning more about their skills and using this to target their job search.

Second, we consider the possibility that the improvements in labor market outcomes from both interventions are driven by only workseeker-side learning. This could occur if workseekers incorrectly believe that firms have limited information. Under this explanation, workseekers in the public certification group believe that firms are more likely to respond to job applications submitted with certificates, hence they change search behavior. This change in search behavior alone, rather than any responses by firms, would then explain the employment and earnings gaps between the public and private certification groups. This seems unlikely, as we observe no differences between these two groups' search targeting or search effort. In the next section, we further address this possibility by presenting an experiment that directly manipulates firms' information, without any scope for changes in workseeker behavior.

[23] We can't reject equality of the public and private certification effects on the offer:application ratio for applications with certificates. This might mean that firms view both certificates as equally credible, although we view this result with caution because the treatment effects on both ratios are very imprecisely estimated and selection into certificate use differs between the two groups.

4.3 Firms Respond to Direct Information Provision

In this section, we show that revealing information only to firms, without allowing any potentially mediating behavior by workseekers, changes their responses to job applications. This is consistent with firms facing information frictions and the employment and earnings effects of certification being partly explained by firm-side responses.

We describe results from an audit-style study here, with more details on the experiment in Appendix E. We invite a random sample of assessed candidates to send us a resume that we will 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 other 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

$$Y_{rv} = \text{Certificate}_{rv} \cdot \beta_1 + \text{Certificate}_{rv} \cdot \text{HighIntensity}_v \cdot \beta_2 + \mathbf{V}_v + \mathbf{X}_r \cdot \Gamma + \mathbf{E}_{rv} + \epsilon_{rv}, \tag{2}$$

where Y_{rv} is the response to resume r sent to vacancy v, Certificate_{rv} is an indicator equal to one if the application includes a public certificate, HighIntensity_v is an indicator equal to one if the vacancy receives 3 applications with certificates rather than 1, $\mathbf{V}_v$ is a vector of vacancy fixed effects that subsumes the main effect of HighIntensity_v, $\mathbf{X}_r$ is a vector of prespecified resume covariates, and $\mathbf{E}_{rv}$ is a vector of fixed effects for the email addresses used to submit the applications. We cluster standard errors by resume and vacancy.[24] We also estimate

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

to explore the vacancy-level effects of getting more applications with certificates. We cannot condition on $(\mathbf{V}_v, \mathbf{X}_r, \mathbf{E}_{rv})$ in the vacancy-level regression but estimates of α are unchanged when we condition on vacancy-level averages of $\mathbf{X}_r$ and sector-of-vacancy fixed effects.

The application-level effects of using a public certificate when other applications do not, captured by β_1, are robustly positive. Applications with a public certificate are 1.6 percentage points more likely to get any response and 1 percentage point more likely to get an interview invitation (Table

[24] 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)
Panel A: Application-level analysis				
Outcome	Any type of response		Interview invitation	
Certificate (β_1)	0.015	0.016	0.009	0.010
	(0.009)	(0.009)	(0.004)	(0.006)
Certificate ×	-0.027	-0.028	-0.016	-0.017
HighIntensity (β_2)	(0.013)	(0.014)	(0.009)	(0.010)
Mean outcome	0.130		0.087	
Vacancy fixed effects		×		×
Email address fixed effects		×		×
Workseeker covariates		×		×
Panel B: Vacancy-level analysis				
Outcome	Response mean	> 0 responses	Invitation mean	> 0 invitations
HighIntensity (α)	0.023	0.042	-0.001	0.021
	(0.020)	(0.026)	(0.016)	(0.021)
Mean outcome	0.134	0.187	0.090	0.117

Note: Analyses in panel A use each of the 3992 applications as an observation. Analysis in panel B use each of the 998 vacancies as an observation. Applications are generated from 717 unique workseekers. Coefficients are from regressing each outcome on a vector of treatment assignments and, in panel A columns 2 and 4, vacancy fixed effects, email address fixed effects, a vector of prespecified workseeker covariates (measured skills, education, age, gender, past employment, and the scan quality of documents they include in their application). The vacancy-level treatment variable HighIntensity$_v$ is included in columns 1 and 3 but omitted in columns 2 and 4 because it is colinear with the vacancy fixed effects. Heteroskedasticity-robust standard errors shown in parentheses, clustered in panel A by resume and vacancy. The mean outcomes in panel A are for applications sent without public certificates to vacancies that receive only one application with a public certificate. The mean outcomes in panel B are for vacancies that receive only one application with a public certificate.

4, panel A, columns 2 and 4). These are substantial effects, both equal to 11% of the control group means, although they are only statistically significant at the 10% level.

These results show that more informative applications lead to higher callback and interview invitation rates in a low-information environment. This suggests firms having limited information plays a role in 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 limited information.

The vacancy-level effects shown in Table 4 panel B are more complex. Vacancies that get more applications with certificates, captured by α, make 2.3 percentage points more callbacks and are 4.2 percentage points more likely to make any callback, although the former effect is not statistically significant and the latter is barely so ($p = 0.099$). The effects on interview invitations are closer to zero.[25]

[25] Vacancies that get more applications with certificates are also significantly more likely to respond to only applications with certificates (5 and 4 p.p. for callbacks and interviews) and less likely to respond to only applications without certificates (2.4 and 0.8 p.p.). The former effect is significantly larger than the latter effect for both applications and interviews. These results show that firms, on average, do not prefer to diversify over applications with and without certificates and do not respond to multiple applications with certificates by becoming suspicious

β_2, the difference between the effect of being the only application with a public certificate sent to a vacancy and the effect of being one of multiple applications with public certificates sent to a vacancy, is negative. Applications that include a public certificate are 2.8 percentage points less likely to get a response and 1.7 percentage points less likely to get an interview invitation when they compete against other applications with certificates (Table 4, panel A, columns 2 and 4). Combining β_1 and β_2 shows that applications with certificates sent to high-intensity vacancies are 0.7-1.2 percentage points less likely to get callbacks and interviews than applications without certificates sent to low-intensity vacancies, although these effects are not statistically significant.

The estimates of β_1, β_2, and α together show that firms respond to more information but that the response may depend on the scale of an information-provision program. A single application containing a public certificate is more likely to get a callback or interview ($\beta_1 > 0$) but this effect shrinks as more applications include public certificates ($\beta_2 \leq 0$), so that the vacancy-level effect of getting more applications with certificates is zero or positive but not statistically significant ($\alpha \geq 0$).

These results on callbacks and interview invitations are consistent with diminishing marginal returns to higher aggregate certificate use, a point we discuss in Section 5.4. However, more informative applications may still be valuable in a higher-information environment for job *offers*, which we do not observe in the audit study. If firms use callbacks and interviews to get more information, then certificates may allow them to interview fewer candidates for each vacancy while still improving match quality and potentially increasing employment.[26]

There are 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 correspondence-based audit studies. 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. These caveats mean that we would need strong assumptions to use the audit results to quantify how much of the public certification effects on employment reflect firm-side responses. Despite these caveats, the audit study does provide additional evidence that firms face limited information.

5 What Do Workseekers and Firms Learn from Skill Certification?

The preceding two sections show that skill certification provides information that improves workseekers' outcomes in the labor market. In this section, we explore what workseekers and firms learn from skill certification, what this implies for the effects of certification for different types of workseekers, and what this might imply for the effects of certification at scale. This section relies on smaller experiments and heterogeneity analysis of the main experiments, so we view these results as more

and discarding all of them.

[26]We do not find robust evidence that outcomes in the high-intensity vacancies depend on the relative skills of the three applications sent with certificates, although this comparison has very low power.

suggestive than conclusive.

5.1 Assessment Results Matter, Not Just Being Assessed

The public certification and audit results above are consistent with three interpretations. First, our preferred interpretation is that firms and workseekers acquire information about workseekers' skills from the assessment results. Second, firms may acquire information about workseekers' tenacity or proactivity from their choice to get assessed, not their actual assessment results. Third, the assessment results may provide no useful information to firms but may be visually appealing or attention-grabbing because they are colorful, branded, and printed on high-quality paper. In this section, we discuss two smaller experiments whose results are consistent with the first but not the second or third interpretations. The first interpretation is also more consistent with the private certification results than the second or third interpretations.

In the first additional experiment, we provide information that workseekers have been assessed without revealing their assessment results. We randomly assign 254 candidates from our workseeker sample, assessed over 3 days, to a 'placebo' certification group. These candidates receive placebo certificates that are identical to the public certificates, including the branding and identifying information, except that the actual assessment results are omitted (Figure F.1) and the psychologist's briefing does not discuss the assessment results.

The placebo certification treatment has minimal effects on labor market outcomes (Table F.1). It increases an index of labor market outcomes by 0.03 standard deviation. This is not significantly different to zero and is significantly smaller than the public certification effect of 0.12 standard deviation. 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 some of the individual outcomes because the small size of the placebo sample leads to large standard errors.

The second additional experiment measures firms' willingness-to-pay (WTP) for information on workseekers' assessment results, conditional on knowing candidates have been assessed. 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.[27] 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

[27] We recruit establishments by asking if they are willing to participate in a study on hiring and tell them we can provide some useful information on hiring. We restrict the sample to establishments that have hiring responsibilities, either single-establishment firms or branches of larger firms that hire independently. 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. See Table G.1 for detailed summary statistics.

resume-style information for our 6,891 candidates. This database allows users to filter and search for candidates with specific assessment results and obtain their contact information. See Figures G.1 and G.2 for selected screenshots of the database. We use a Becker-DeGroot-Marschak mechanism to measure WTP for access to this database relative to a placebo database with candidates' contact information and selected resume-style information, but no assessment results (Becker et al., 1964).[28]

Firms' WTP for access to the database with assessment results is substantial: 68% of firms report positive WTP and the unconditional mean WTP is 1,161 South African rand or USD 195 PPP (Figure G.3). Mean WTP is 224% of the mean weekly earnings for employed candidates in our workseeker sample. This shows firms value information on specific assessment results, conditional on knowing candidates have been assessed.

Both the placebo experiment and WTP measurement are consistent with the first but not second or third interpretations above: information about assessment results is valuable, not just information about whether candidates have been assessed or any visual appeal of the certificates. This provides additional support for our preferred interpretation: public certification provides information about workseekers' types and either facilitates more productive firm-worker matches or lowers screening costs.

5.2 Horizontal Versus Vertical Differentiation of Workseekers

Public certification provides more information about workseekers' types, allowing these types to be differentiated more accurately. Our conceptual framework distinguishes two types of workseeker differentiation. Under horizontal differentiation, type i workseekers are more productive than type j workseekers in type i jobs, and vice versa. Under vertical differentiation, type i workseekers are more productive than type j workseekers in both type i and j jobs. Under horizontal differentiation, additional information can help both types of workseekers by matching them with jobs where they are more productive. Under vertical differentiation, additional information can hurt type j workseekers by reducing their probability of being mistaken for more productive type i workseekers. Our experiments are not primarily designed to test vertical versus horizontal differentiation but we present some suggestive evidence on this distinction.

We observe two patterns in our data that are not consistent with at least some models of vertical differentiation. First, the public certification effects of employment are not robustly increasing in measured skill. To show this, we construct three indices that combine the six assessment results in different ways: the number of top terciles minus bottom terciles, the first principal component

[28] We ask how much they are willing to pay for three months of database access, and then randomly offer them a discount between 0 and 100% of the normal price of 10,000 South African Rands (USD 1,670 PPP). 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 explain the entire mechanism and run a practice round before the official round. See Appendix G for more details on the experimental protocol.

of the cardinal scores, and a weighted average of the cardinal scores with weights based on their associations with earnings.[29] 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 include this indicator and its interactions with treatment assignments in equation (1). The interaction effects with public certification on employment are smaller than 2 percentage points and not significantly different to zero for all indices (Table D.11 panel A).[30]

Second, public certification does not increase the dispersion of earnings conditional on employment. To show this, we estimate the standard deviation, interquartile range, and interdecile range of earnings conditional on employment in the public certification and control groups. These estimates are respectively 0.03, 0.65, and 0.42 inverse hyperbolic sine points lower in the public certification group than the control group. The latter two differences are substantial but none are close to statistically significant using a clustered nonparametric bootstrap test (p = 0.87, 0.57, and 0.41 respectively). This pattern is inconsistent with one form of vertical differentiation, where workers have a single index of skill, productivity is monotonically increasing in skill, skill is observed with classical measurement error, and workseekers are hired only if their imperfectly observed skill exceeds some threshold. In this model, public certification would increase the dispersion of earnings conditional on employment through two mechanisms. For inframarginal workers who are employed with or without certification, certification would steepen the earnings-skill gradient, raising the earnings dispersion. Marginal workers who are employed only with certification will be close to the bottom of the earnings distribution, hence raising the dispersion of earnings conditional on employment. Neither dispersion tests we report here nor the quantile treatment effects we report in Section 3.3 match the predictions of this model of vertical differentiation.

Why do we see little evidence of vertical differentiation in this setting? We document three mechanisms that can lead to more horizontal than vertical differentiation in this setting.

First, there is substantial heterogeneity in firms' relative demand for different skills.[31] 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

[29] The weights equal the coefficients from regressing earnings on the cardinal scores using control group data. Results are similar for weighted averages based on the coefficients of regressions of control group earnings on polynomial or spline functions of the skills.

[30] We see similarly little evidence of heterogeneous treatment effects by skill when we use continuous indices instead of binary indicators and when we use alternative model specifications: using nonlinear functions of skill indices that allow non-monotonic relationships, using different single indices, or using machine learning methods to estimate heterogeneous treatment effects simultaneously across all individual scores.

[31] We also estimate earnings-skill gradients in the control group of workseekers and compare these across skills. These are relatively similar for all skills except communication, which has a slightly steeper gradient than the others. This is consistent with different types of firms valuing different skills. But we view this as weak evidence, because the estimated earnings-skill relationships condition on endogenous firm-worker matching.

workseekers from the online database. 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: the share of firms ranking each profile highest ranges from 6 to 33%. The seventh profile has middle terciles for all six assessments and has a one-year post-secondary education 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. We find similar results when we ask firms to rank profiles with visible versus concealed assessment results. See Appendix G for more details on the experimental protocol and results.

Second, assessment results are weakly correlated across skills within candidates. 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 several pairs of skills (Table A.2). As a result, most candidates' certificates show substantial variation across skills. Table A.3 shows that 88% of the candidates have at least one top tercile but only 24% have four or more top terciles and only 2.3% have all top terciles. 76% of the candidates have at least one bottom tercile but only 12% have four or more bottom terciles and only 0.7% have all bottom terciles. 64% of candidates have both top and bottom terciles. Other studies that measure multidimensional skills also find weak correlations across skills within candidates (Almlund et al., 2011; Poropat, 2009).

Third, workseekers with different skills respond differently to public certification. To show this, we regress both search targeting and certificate use with job applications on the same treatment $\times$ skill index interactions described earlier in this subsection. Workseekers with relatively high skills are more likely to use certificates in job applications. Workseekers with relatively low skills are more likely to engage in search targeting, although this difference is not statistically significant.

These three mechanisms show how public certification can facilitate horizontal more than vertical differentiation in this setting. Different firms demand different skills, different workseekers supply different combinations of skills, certification helps workseekers target jobs that value their skills, and certification helps firms hire workseekers whose skills better match the firms' demand. These patterns are consistent with models of multidimensional skill where information frictions can lead to poor matches between workseeker skills and firm requirements (Fredriksson et al., 2018; Guvenen et al., 2020; Lise and Postel-Vinay, 2020).

However, our experiments are not primarily designed to test horizontal against vertical differentiation, so we view this as suggestive evidence that can motivate future work. Certification may facilitate vertical rather than horizontal differentiation when it is based on assessment of a single skill or when certificates show only a single summary measure of multiple skills, unlike our approach of measuring and reporting multiple weakly correlated skills. Certification may also facilitate vertical rather than horizontal differentiation when it covers a larger share of the workforce, whereas our sample excludes highly educated and highly experienced workseekers.

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. Public certification effects on employment are 2.7 percentage points smaller for candidates with post-secondary education (standard error 2.8 p.p.) and 4.3 percentage points smaller for candidates with prior work experience (standard error 3.2 p.p.) (Table D.11, panel B). We also estimate the latent probability of being employed at endline as a single summary measure.[32] Candidates with above-median latent probabilities of employment have a 7.6 percentage point smaller public certification effect than candidates with below-median latent probabilities (standard error 2.8 p.p.). These results show that certification can substitute for traditional sources of information about workseekers' skills.[33] This is consistent with evidence that educational qualifications are more useful for members of groups facing statistical discrimination (Arcidiacono et al., 2010).

5.4 Skill Certification at Different Scales

We show that skill certification at a relatively small scale increases employment and earnings for certified workseekers. In this section we discuss conditions under which effects may vary with the scale of skill certification. This provides a guide for thinking about potential scale effects, rather than a confident or quantitative argument about scale effects.[34]

First, employment and earnings effects may depend on scale if certified workseekers displace non-certified workseekers. It is unlikely that our experimental results are due to displacement of non-certified workseekers *in the control group*. We certify only 2,247 workseekers in a metropolitan area with roughly 8 million people and 2 million employed workers (Statistics South Africa, 2016b). The probability of certified and control group workseekers applying for the same jobs by chance

[32] 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, adjusting the predicted values in the control group using leave-one-out estimation to avoid overfitting.

[33] This result is not explained by a correlation between workseekers' skills and their education and past employment. We regress employment on treatment assignments, a single index measure of skill from Section 5.2, a measure of information about workseekers' skills from this section, and a full set of interactions. The interactions between public certification and the single index skill measure remain close to zero, while the interactions between public certification and the measure of information about workseekers' skills remain negative.

[34] There are few existing papers that study how the effects of specific active labor market policies change with scale. For job search assistance policies specifically, Crépon et al. (2013) and Lise et al. (2004) find larger-scale policies generate negative spillovers on non-participants, while Blundell et al. (2004) find no spillovers on non-participants.

is very small, and Harambee does not encourage recently-assessed workseekers to apply to specific jobs or search for work in specific areas.

It is possible that certified workseekers displace non-certified workseekers who are not part of the experimental sample. We cannot directly test for this, but we can evaluate the mechanisms that might generate it. Displacement is less likely if certification improves match quality or reduces screening costs and hence increases the share of latent vacancies that are worth filling, as in our conceptual framework and general equilibrium models of information frictions (Jovanovic, 1979; Gonzalez and Shi, 2010). Displacement is more likely if firms value certification for some reason other than information (e.g. visual appeal) or if certification helps firms to identify a small set of universally-demanded workseekers and compete for them.

Our results are more consistent with the match quality or screening costs mechanisms. We find that firms' demand for different skills is heterogeneous, firms value learning about workseekers' specific skill types, and the gains from certification are not limited to workseekers with specific skill profiles. All these patterns suggest that firms and workseekers use certification to learn about workseekers' skills and achieve some combination of better matches between workseekers' skills and firms' demand or equally good matches at lower screening cost. We also find that certification increases earnings and hourly wages conditional on employment, suggesting that certified workseekers are in matches that generate more value net of screening costs. We do find that the callback and interview premia to certification drop when certified applicants compete against each other in the audit study. This is consistent with some certified workseekers displacing other certified workseekers from interviews. However, as discussed in Section 4.3, this does not necessarily imply that certified workseekers displace other certified workseekers from hiring. Certification may allow firms to call back and interview fewer candidates for each vacancy and still make better-matched hires. This suggests one specific way certification can reduce screening costs and is consistent with the finding by Algan et al. (2020) that reducing firm-level screening costs can raise hiring.

Second, employment and earnings effects may depend on scale if the extent of limited information varies across the population of either firms or workseekers. Consider the case where the population is divided into fraction p of uninformed workseekers, who do not know their skills and cannot convey their skills to firms, and fraction $1 - p$ of informed workseekers, who know their skills and can convey this information to firms. Assessing and certifying the latter group will have limited returns. Our finding that certification has larger employment and earnings effects when there are limited alternative sources of information on workseekers' skills is consistent with this possibility. Our experiment does not identify the population shares of workseekers or firms facing information frictions. The share of relatively uninformed types may be higher in our sample than the population, as we study workseekers with poor baseline labor market outcomes. But Harambee's workseeker recruitment does not explicitly mention assessments or information frictions, so workseekers are unlikely to select into the sample specifically for assessment and certification.

Third, employment and earnings effects may depend on scale if certificate (non-)use conveys information in general equilibrium. If, for example, all workseekers get assessed and certified but only use certificates when applying to vacancies where their match quality is high, then firms may infer that workseekers without certificates are poor matches for these vacancies. In another example, some firms may choose not to use assessments in hiring if assessments are costly and they believe they can infer workseekers' types by observing their interactions with other firms (Lockwood, 1991). Our experiments cannot speak to these general equilibrium mechanisms. But adding either of these two mechanisms to our conceptual framework still predicts that any non-zero use of assessment and certification will raise employment and earnings relative to no assessment and certification.

Even if reducing information frictions has decreasing effects on employment and earnings at larger scales, it may still raise workseeker or firm welfare by reducing job search costs, vacancy posting costs, and the frequency of bad hires that lead to separations. This interpretation is consistent with models showing that firm- and workseeker-level search and matching frictions, including information frictions, can lower aggregate utility through multiple mechanisms, not just through unemployment (Donovan et al., 2018; Mortensen and Pissarides, 1999; Poschke, 2019).

6 Conclusion

We find that workseekers make different job search decisions, firms make different interview decisions, and workseekers experience higher employment and earnings when more information is available about workseekers' skills. Assessing workseekers' skills and communicating the assessment results to both workseekers and firms increases assessed workseekers' employment by 17% (5 percentage points), earnings by 34%, and hourly wages by 20%. This shows that skill certification gets more workseekers into jobs and that these jobs pay more. When workseekers learn their assessment results but cannot easily and credibly share assessment results with firms, their labor market outcomes improve, but not by as much. This shows the importance of getting credible information to both sides of the market.

We study a context and sample where information frictions are likely: work experience is limited, education-skill relationships are relatively weak, hiring mistakes are costly, and reservation and minimum wages are relevant. However, none of these features is 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 that, in similar contexts, providing information about workseekers' skills may be a valuable focus of government policy. Some existing job search assistance programs offer skill assessments to workseekers (McCall et al., 2016). Adding certification to these assessments might enhance their effectiveness at low cost. We find that adding certification to an existing assessment program generates earnings gains for workseekers that easily exceed the cost of both assessment and certification. Government involvement, through public-sector assessment programs or subsidies to private-sector assessments, is likely to be particularly important for credit-constrained workseekers (Abebe et al., 2020b). Better information about workseekers' skills could also come from more accurate assessments during formal education (MacLeod et al., 2017).

Our results suggest there may also be scope for market-based provision of information about workseekers' skills. We show that firms are willing to pay for access to 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 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 show that providing more information through certification can be valuable even in a labor market where some firms already use assessments, suggesting 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 can establish reputations, how precisely or coarsely information should be reported, and under what conditions participants will opt into or out of assessment. This work might incorporate existing models of screening and signalling when both agents and principals have limited information, allowing possible interaction effects (Alonso, 2018; Rosar and Schulte, 2012).

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.[35]

[35] We find 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,

Finally, our results show that certification allows some combination of higher match quality and lower screening costs for firms. Quantifying the relative importance of these mechanisms is difficult without direct data on firm recruitment practices and productivity. Future work could explore this further, by combining data on both earnings and productivity (as in Kahn and Lange 2014) with variation in firms' information about workseekers' skills.

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 (2020a): "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City," *Review of Economic Studies*, forthcoming.

Abebe, G., S. Caria, and E. Ortiz-Ospina (2020b): "The Selection of Talent: Experimental and Structural Evidence from Ethiopia," Manuscript, University of Bristol.

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 (2020): "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.

Algan, Y., B. Crepon, and D. Glover (2020): "Are Active Labor Market Policies Directed at Firms Effective? Evidence from a Randomized Evaluation with Local Employment Agencies," Manuscript, Sciences Po.

Almlund, M., A. Duckworth, J. Heckman, and T. Kautz (2011): "Personality Psychology and Economics," in *Handbook of the Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessmann, Elsevier, 1–181.

Alonso, R. (2018): "Recruiting and Selecting for Fit," Manuscript, LSE.

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.

rather than a core result.

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

Anderson, M. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481–1495.

Arcidiacono, 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.

Banerjee, A. and G. Chiplunkar (2020): "How Important Are Matching Frictions in The Labor Market? Experimental & Non-experimental Evidence from a Large Indian Firm," Manuscript, MIT.

Banerjee, A. and S. Sequeira (2020): "Spatial Mismatches and Imperfect Information in the Job Search," Discussion Paper 14414, Centre for Economic Policy Research.

Bassi, V. and A. Nansamba (2020): "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 Labor 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.

Becker, G., M. Degroot, and J. Marschak (1964): "Measuring Utility by a Single-response Sequential Method," *Behavioral Science*, 9, 226–232.

Belot, M., P. Kircher, and P. Muller (2018): "Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice," *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.

Bryan, G., S. Choudhury, and A. M. Mobarak (2014): "Under-Investment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 82, 1671–1748.

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.

Chandrasekhar, A. G., M. Morten, and A. Peter (2020): "Network-Based Hiring: Local Benefits; Global Costs," Working Paper 26806, National Bureau of Economic Research.

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.

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

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.

Franklin, S. (2017): "Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies," *The Economic Journal*, 128, 2353–2379.

Fredriksson, P., L. Hensvik, and O. N. Skans (2018): "Mismatch of Talent: Evidence on Match Quality, Entry Wages, and Job Mobility," *American Economic Review*, 108, 3303–3338.

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.

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

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

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2020): "Multidimensional Skill Mismatch," *American Economic Journal: Macroeconomics*, 12, 210–244.

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

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

Ingle, K. and C. Mlatsheni (2017): "The Extent of Churn in the South African Youth Labour Market: Evidence from NIDS 2008-2015," Manuscript, University of Cape Town: Southern Africa Labour and Development Research Unit.

International Labour Organization (2016): *Enabling Environment for Sustainable Enterprises in South Africa*, Geneva: International Labour Office.

——— (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.

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.

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.

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.

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.

Lockwood, B. (1991): "Information Externalities in the Labour Market and the Duration of Unemployment," *Review of Economic Studies*, 58, 733–753.

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.

Malindi, K. (2017): "Imperfect Information and The Racial Wage Gap for South African Men," Manuscript, University of Stellenbosch.

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.

Mortensen, D. and C. Pissarides (1999): "New Developments in Models of Search in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, 2567–2627.

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.

Poropat, A. E. (2009): "A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance," *Psychological Bulletin*, 135, 322–338.

Poschke, M. (2019): "Wage Employment, Unemployment and Self-employment across Countries," Manuscript, McGill University.

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

Pugatch, T. (2018): "Bumpy Rides: School-to-work Transitions in South Africa," *Labour*, 38, 205–242.

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

Rosar, F. and E. Schulte (2012): "Optimal Test Design under Imperfect Private Information and Voluntary Participation," Manuscript, University of Bonn.

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

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.

Statistics South Africa (2016a): "Labour Market Dynamics in South Africa 2018 [dataset]. Version 1," Pretoria: Statistics South Africa [producer], 2017. Cape Town: DataFirst [distributor], 2018. DOI: https://doi.org/10.25828/kkzz-mr59.

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

——— (2017): "Labour Market Dynamics in South Africa 2018 [dataset]. Version 1," Pretoria: Statistics South Africa [producer], 2018. Cape Town: DataFirst [distributor], 2019. DOI: https://doi.org/10.25828/wyv8-6k31.

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?" Working Paper, Stellenbosch University.

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.

Online Appendix
"Job Search and Hiring with Limited Information about Workseekers' Skills"
Eliana Carranza, Robert Garlick, Kate Orkin, Neil Rankin

A Assessments

We assess each workseeker's skills in six domains. 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 the 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 settings.

A.1 Firms' Use of Assessments

Harambee has used the numeracy, communication, and concept formation assessments since 2011 to select candidates for further job readiness training and recommend candidates to vacancies at partner firms. Harambee has placed over 160,000 candidates in entry-level jobs using these assessments. Table A.1 shows how 33 large client firms in retail, hospitality, logistics and corporate services require Harambee to use assessments when recommending candidates for interviews.

All firms used at least one assessment to screen candidates and 73% of firms used all three assessments. In contrast, only 57% required certified results on the national high school graduation exam and only 3% required references. This shows firms find this skill information useful relative to other sources of information about prospective workers' skills. 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 could not include this assessment in the certification because it is a proprietary instrument.

We therefore selected three alternative measures of skills which would be unlikely to be correlated with numeracy, communication, and concept formation. To select these, we conducted interviews with 20 hiring managers to understand which other skills they valued in successful hires. Elsewhere, we conducted a detailed literature review of measures and selected those most overlapping with what firms valued (Esopo et al., 2018), which were also correlated with either earnings or measures of workplace performance in some settings.

A.2 Description of Assessments

Concept formation is very similar to the Raven's Progressive Coloured Matrices assessment (Raven and Raven, 2003). It is a non-verbal measure of fluid intelligence, which captures the rate at which people learn and their conceptual reasoning. It specifically assesses the ability to ignore superficial differences and see underlying commonalities across situations and to use logic in new situations. Meta-analyses identify measures of fluid intelligence as strong predictors of worker

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

Sector	# firms	% of firms using each piece of information to screen candidates						
		Assessment result for			Career aptitude profile	Criminal record check	High school graduation certificate	Reference
		Communi-cation	Concept formation	Numeracy				
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 are coded as using an assessment if they require candidates to reach a certain threshold score on the assessment to be eligible for interviews or training programs. Firms are coded as using other documents if they require these to be submitted with the candidates' application packages. The criminal record 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 Harambee for candidates whom Harambee shortlists for interview, not how firms use the information. Data are from direct conversation with Harambee staff.

productivity (Schmidt and Hunter, 1998; Schmidt et al., 2016). The Raven's test is widely used in hiring and selection (Chamorro-Premuzic and Furnham, 2010), including in recent research in economics (Abebe et al., 2020b; Beaman et al., 2018). Scores on this assessment are correlated with interview ratings, technical scores and supervisor ratings in several South African firms (De Kock and Schlechter, 2009; Lopes et al., 2001; Taylor, 2013).

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 (www.mediaworks.co.za) and assesses proficiency in arithmetic used in high school: comparing different types of numbers; working with fractions, ratios, money, percentages and units; and performing calculations with time and area.

Communication captures English language listening, reading and comprehension skills. The assessment was developed by a South African adult education provider (www.mediaworks.co.za) and is designed to assess English proficiency for high school students. It evaluates both listening and written comprehension. It focuses on ability to identify and recall the main message of a text or passage, infer meaning of vocabulary through context clues, and infer meaning when information is not directly stated. Both numeracy and communication skills are correlated with educational attainment and wages in OECD countries (Heckman et al., 2006; Heckman and Kautz, 2012; Hanushek et al., 2015). There are also correlations between wages and numeracy (du Rand et al., 2011) and wages and English communication skills (Casale and Posel, 2011) in South Africa, conditional on education.

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 is a validated self-reported 8-item psychological scale (Duckworth et al., 2007). Grit correlates with academic performance and workplace retention in the US (Eskreis-Winkler et al., 2014).

The assessment labeled *Focus* on certificates captures inhibitory control, the ability to distinguish relevant from irrelevant information, control one's attention to focus on what is needed for a task (Diamond, 2013) and guide thought and action in accordance with a goal (Posner and DiGirolamo, 1998). The assessment is a computerized version of the widely-used Stroop Test, using colors (Stroop, 1935). Similar measures are correlated with employment status (Kalechstein et al., 2003) and 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 the Hit 15 lab task (Gneezy et al., 2010). The computer and the subject take turns adding either one, two or three points to the points basket. The goal is to be the first player to reach 15 points. It captures ability to search for relevant information and anticipate the consequences of actions. High planning scores predict retention rates among truckers in the US, conditional on cognitive skills (Burks et al., 2009). Similar measures of complex planning skills are correlated with wages in South Africa, controlling for fluid intelligence and education (Ederer et al., 2015).

For the first 17 of the 84 assessment days, covering 26% of candidates, computer problems meant that we used two self-reported psychological scales, labeled *Control* and *Flexibility* on the certificates instead of focus and planning. We used two subscales of the Personal Problem-Solving Inventory (Hepner and Petersen, 1982). The Personal Control scale (control) captures whether candidates take a systematic or impulsive and erratic approach when faced with new, challenging problems. The Approach Avoidance (flexibility) scale captures whether candidates actively consider several approaches to solving a problem or whether they pursue their first idea without thinking about alternatives. These are not exact analogues of the tasks: they capture self-perceptions as well as behaviors (Heppner, 1988). But scores are correlated in other samples: for example, the PSI is correlated with the Stroop task (Rath et al., 2004). 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.

We use the assessment scores 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 similar when

Table A.2: Correlations of Assessment Results

	Concept formation	Grit	Numeracy	Control	Flexibility
Panel A: Correlations In First 17 Days of Assessment (1615 workseekers)					
Communication	0.337	0.127	0.386	0.237	0.126
Concept formation		0.108	0.489	0.174	0.098
Grit			0.162	0.507	0.334
Numeracy				0.212	0.107
Control					0.173
Panel B: Correlations In Remaining 67 Days of Assessment (5276 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			0.128	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. 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 A.3: Distribution of Top, Middle, and Bottom Terciles Shown on Candidates' Reports

		Fraction with _ bottom terciles							Total
		0	1	2	3	4	5	6	
Fraction with _ top terciles	0	0.001	0.007	0.025	0.032	0.029	0.018	0.007	0.119
	1	0.009	0.036	0.059	0.064	0.037	0.011	-	0.215
	2	0.027	0.077	0.079	0.040	0.011	-	-	0.235
	3	0.054	0.076	0.048	0.009	-	-	-	0.187
	4	0.070	0.059	0.009	-	-	-	-	0.138
	5	0.060	0.024	-	-	-	-	-	0.084
	6	0.023	-	-	-	-	-	-	0.023
Total		0.243	0.279	0.220	0.146	0.076	0.029	0.007	

Table shows the share of the sample with i top terciles and j top terciles on their reports for each $i, j \in \{0, 6\}$. The number of middle terciles equals $6 - i - j$.

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 that each candidate took only six of the eight assessments. The profile-ranking exercise does not use the control or flexibility scales.

A.3 Administration of Assessments

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 to their assessment results being shared with Harambee, the research team, and external firms.

Registered industrial psychologists employed or contracted by Harambee oversaw administration of all assessments. They also delivered briefings to candidates to interpret results. Finally, the lead psychologist at Harambee approved the language on certificates. This ensures compliance with South African law on psychometric testing in workplace settings.

A.4 Validation of Self-Reported Psychological Scales and Tasks

We use four self-reported psychological scales in the paper: grit, control and flexibility are used as skills measures, while self-esteem is used as an outcome measure. We followed standard procedures in psychology to ensure the self-reported scales were well-understood and valid as measures. See Esopo et al. (2018) for a full discussion of the process followed. We use the same seven-point Likert scale for all scales.

The Problem-Solving Inventory had already been validated in South Africa with young black African students of a very similar demographic profile to our sample and we used this item wording (Pretorius, 1993; Heppner et al., 2002). For grit and self-esteem, we ensured language used was well-understood by conducting 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?" We simplified the wording of some items and altered some culturally specific idioms in response to the cognitive debriefings.

Second, we estimated the extent to which different items in each scale move together, using Cronbach's alpha (Cronbach, 1951). All assessments have $\alpha > 0.65$. Third, we administered the scales twice for 150 candidates, ten days apart. We estimated Lin's Concordance Correlation

Coefficient (Lawrence and Lin, 1989) between the two administrations. All assessments have $\rho_c > 0.62$. Fourth, we check if any items on the scales have very low variation across candidates using maximum endorsement frequencies. No items meet the threshold for being dropped due to insufficient variation from Bowling (2014).

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.

Table A.2 shows the correlation of assessment results for the different skills. Numeracy, concept formation and communication have pairwise correlations of 0.34 to 0.52. Numeracy and communication assessments capture acquired knowledge, often from schooling, which is often positively correlated with fluid intelligence. This is potentially because learning at a higher rate improves acquisition of knowledge (Heckman and Kautz, 2012; Nisbett, 2009; Roberts et al., 2000). However, as we intended, these are less strong correlations between the other tasks (focus and planning) and the scales (grit, flexibility, and planning). These suggest the certificates will horizontally differentiate workseekers from one another.

B Implementation Costs

This appendix reports the costs of the public certification intervention and compares these to gains experienced by treated workseekers, showing that the latter easily exceed the former. We measure costs from the Harambee and J-PAL Africa financial statements. All costs are reported in 2016/7 PPP USD terms and are averaged over the 2,247 candidates who received the public certification intervention. The cost figures in nominal USD are 42% of the cost figures in PPP USD, though this does not affect the cost-benefit comparisons. We report average variable costs and, where these are possible, total and average fixed costs. The average variable costs may change with scale but we do not attempt to project scale effects on costs.

The average variable cost of adding certification to Harambee's existing assessment operation was USD 23.10. This included certificate printing, software license fees, website hosting fees, the time of J-PAL and Harambee staff used to prepare the certificates, and the time of Harambee psychologists used to conduct briefings. This also included a USD 10.32 transport subsidy to each participant to cover the cost of travel to the Harambee office, which is arguably not a necessary cost of the intervention. These cost calculations exclude the private and placebo certifications, audit study, and firm-facing experiments.

The average variable cost of certification and assessment was USD 57.27 per participation. This included all certification-only costs, facility rental, computer rental, data and internet costs, and the time of Harambee staff who administered the assessments. Facility and computer rental costs were the largest line items for the assessment cost, jointly accounting for USD 23.43.

The average variable costs exclude fixed costs such as licenses for the assessment tools, market research into firm preferences over assessments, and senior management fees. For these costs we either cannot calculate a meaningful average fixed cost or cannot reliably separate Harambee's total fixed costs for developing the assessment program from its costs of other activities. J-PAL Africa's fixed cost for developing the certification program on top of the assessment program was approximately USD 17,685 or USD 7.87 per candidate who received the public certification intervention. This covered J-PAL Africa staff costs during development and all costs of piloting the certificates with firms and workseekers. This includes the cost of developing and piloting the private and placebo certifications, which we cannot easily separate from the public certification, but excludes the costs of developing and piloting the audit study and firm-facing experiments.

We compare these average costs to the average benefit per participant who received the public certification intervention over the first three months after the intervention. Public certification increases average earnings by USD 9.05 in the week before the endline survey and the endline survey occurred on average 14.4 weeks after treatment. Multiplying these together gives an average effect on earnings since treatment of USD 130.2: 5.6 times higher than the average variable cost of certification, 2.3 times higher than the average variable cost of assessment and certification, and 2.0 times higher than the average variable cost of assessment and average variable and fixed costs of certification. The gains to treated workseekers over just three and a half months easily exceed the cost of public certification and assessment.

The preceding calculation assumes that the treatment effect on weekly earnings does not vary through time from treatment to the endline. The public certification effect on earnings does not substantially vary with the time period from treatment to endline. But the treatment effects on recalled employment in the first and second months after treatment are not identical, suggesting a possible time trend (Table D.12). To account for this, we convert the weekly earnings effect into monthly terms and multiply this by the sum of the employment effect in the first month after treatment, the second month after treatment, and the week before the endline. This gives an average on earnings since treatment of USD 110.1, which also easily exceeds the cost of public certification and assessment.

C 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. We use the term "treatment" to refer to the public certification. 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:

$$
\begin{aligned}
&\underbrace{\mathbb{E}[Earn|Treat=1]-\mathbb{E}[Earn|Treat=0]}_{\text{ATE for earnings}} \qquad (4)\\
=\;&\underbrace{\left(\mathbb{E}[Earn|Treat=1,Work=1]-\mathbb{E}[Earn|Treat=0,Work=1]\right)}_{\text{ATE for earnings | employment}}\cdot\underbrace{Pr[Work=1|Treat=1]}_{\text{Treated employment rate}}\\
+\;&\underbrace{\mathbb{E}[Earn|Treat=0,Work=1]}_{\text{Control earnings | employment}}\cdot\underbrace{\left(Pr[Work=1|Treat=1]-Pr[Work=1|Treat=0]\right)}_{\text{ATE for employment}}.
\end{aligned}
$$

We define the second line on the right-hand of the regression as the extensive margin effect. Intuitively, this is the average treatment effect on employment 'priced' at the mean earnings value in the control group. If treatment has no effect on the employment rate, then this expression is zero. We define the first line on the right-hand side of the regression as the intensive margin effect. If treatment only changes the employment rate but has no effect on earnings for employed candidates, then this term is zero.[36]

All terms in equation (4) 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 (4). 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 modeling the 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 in our setting: the former methods require an instrument for selection into employment that we do not have and the

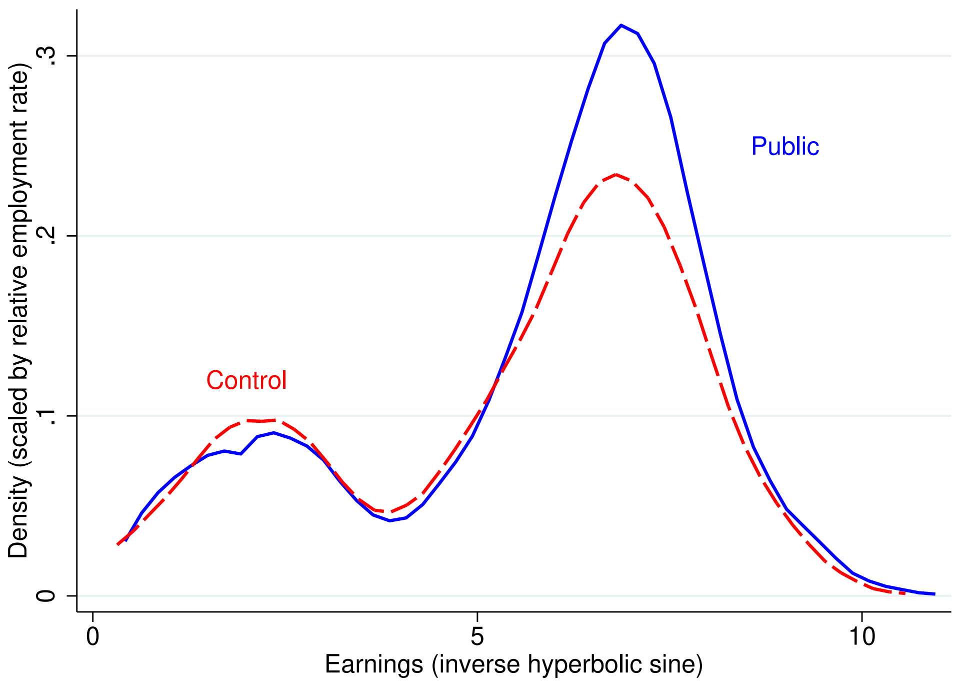
[36]Attanasio et al. (2011) show that the intensive margin effect can be further decomposed into two terms: the treatment effect on earnings conditional on candidates' baseline characteristics, and the difference in baseline characteristics between employed candidates in the treatment and control groups. However, neither of these terms is point identified. Separating these effects is not important in our application. Our conceptual framework is consistent with certification either increasing the same workseekers' latent treated wages conditional on employment, or increasing mean wages conditional on employment by helping workseekers with higher latent treated wages get employed.

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 and wage 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 (in the main text) shows the distributions of earnings for each group and the quantile treatment effects of public certification. Figure C.1 shows 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, the vertical difference between the densities at each earnings level E represents the treatment effect on the share of all candidates earning E, not on the share of employed candidates earning E. The treatment effect on the earnings density is almost entirely above median earnings for employed control group candidates. This shows that either the marginal candidates employed only when treated earn more than most inframarginal control candidates, or treatment increases earnings for inframarginal candidates, or both.

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



This figure shows the densities of earnings in the control and public certification 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. Hence the vertical difference between the densities at each earnings level E represents the treatment effect on the share of all candidates earning E, not on the share of employed candidates earning E. The density is estimated only for the employed, so candidates with zero earnings are excluded.

D Additional Results about Workseeker Experiments

D.1 Summary Statistics and Balance Tests

This section reports summary statistics for the baseline workseeker sample (Table D.1) and endline workseeker sample (Table D.2). Table D.3 assesses balance in the baselined and endlined samples by showing group-specific means and p-values for tests for equal means. Balance tests for equal means of baseline measures are also reported in the final column of Table D.1. Table D.4 compares our workseeker sample to the broader population of the country and of Gauteng province, where the study took place.

Table D.1: Summary Statistics for Baseline Variables

Variable	# obs	Mean	Std dev.	10^{th} pctile	90^{th} pctile
Age	6891	23.6	3.3	19.8	28.3
Male	6891	0.382	0.486		
University degree / diploma	6891	0.167	0.373		
Any other post-secondary qualification	6891	0.212	0.409		
Completed secondary education only	6891	0.610	0.488		
Panel B: Assessment Results					
Numeracy score	6891	0.000	1.000	-1.253	1.376
Communication score	6891	0.000	1.000	-1.152	1.656
Concept formation score	6891	0.000	1.000	-1.577	1.224
Grit score	6891	0.000	1.000	-1.354	1.259
Other scores	6701	0.000	1.086	-1.340	1.324
Panel C: Labor Market Measures					
Employed	6891	0.378	0.485		
Earnings	2116	565	740	100	1400
Ever worked	6877	0.704	0.457		
Ever held a long-term job	6877	0.090	0.286		
Panel D: Job Search Measures					
Searched	6891	0.968	0.175		
Applications submitted[a]	6815	9.9	18.6	2.0	20.0
Search cost	6147	242	1520	30	400
Search hours	6699	17.0	20.8	2.0	48.0
Offers received[a]	6810	1.20	7.20	0.00	2.00
Panel E: Belief Measures					
Planned applications[a]	6840	48.9	1629.9	4.0	36.0
Correct about all assessment results	6891	0.082	0.274		
Incorrect about all assessment results	6891	0.290	0.454		
Overconfident about all assessment results	6891	0.219	0.413		
Underconfident about all assessment results	6891	0.010	0.100		

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 ≈ USD 0.167 in purchasing power parity terms. Intensive-margin labor market measures (e.g. earnings) are set to missing for non-workers. Intensive-margin search measures (e.g. search cost) 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).

Table D.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 ≈ USD 0.167 in purchasing power parity terms. Intensive-margin labor market measures (e.g. earnings) are set to missing for non-workers. Intensive-margin search measures (e.g. search cost) 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).

Table D.3: Balance Tests in Baseline and Endline Samples

	Baseline Sample Means				Endline Sample Means			
Variable	Control	Private	Public	p:equal	Control	Private	Public	p:equal
Age	23.5	23.8	23.7	0.583	23.5	23.8	23.7	0.546
Male	0.389	0.365	0.387	0.267	0.386	0.360	0.385	0.216
University degree / diploma	0.158	0.178	0.171	0.889	0.151	0.175	0.169	0.819
Any other post-secondary qualification	0.214	0.223	0.202	0.642	0.217	0.227	0.204	0.599
Completed secondary education only	0.617	0.593	0.612	0.794	0.620	0.592	0.612	0.752
Panel B: Assessment Results								
Numeracy score	-0.002	-0.018	0.024	0.523	-0.007	-0.019	0.016	0.672
Communication score	0.038	-0.002	-0.029	0.206	0.031	-0.004	-0.032	0.237
Concept formation score	0.020	-0.012	-0.005	0.764	0.017	-0.017	-0.014	0.747
Grit score	-0.045	0.026	0.018	0.089	-0.042	0.028	0.028	0.096
Other scores	0.020	-0.010	-0.003	0.851	0.024	-0.015	-0.005	0.763
Panel C: Labor Market Measures								
Employed	0.364	0.386	0.387	0.468	0.362	0.388	0.386	0.434
Earnings	609	584	517	0.083	607	582	513	0.064
Ever worked	0.693	0.716	0.703	0.418	0.690	0.713	0.706	0.397
Ever held a long-term job	0.095	0.090	0.086	0.696	0.095	0.086	0.087	0.571
Panel D: Job Search Measures								
Searched	0.967	0.975	0.960	0.058	0.967	0.977	0.960	0.028
Applications submitted[a]	9.9	10.1	9.6	0.809	9.6	10.0	9.7	0.892
Search cost	205	240	280	0.276	205	243	285	0.258
Search hours	17.6	17.0	16.4	0.231	17.5	16.8	16.4	0.262
Offers received[a]	1.00	1.41	1.12	0.280	1.02	1.46	1.15	0.241
Panel E: Belief Measures								
Planned applications[a]	19.7	22.4	107.0	0.252	19.6	22.5	110.7	0.247
Correct about all assessment results	0.083	0.081	0.083	0.960	0.083	0.080	0.082	0.944
Incorrect about all assessment results	0.287	0.291	0.291	0.961	0.288	0.292	0.291	0.972
Overconfident about all assessment results	0.216	0.215	0.225	0.732	0.216	0.216	0.226	0.714
Underconfident about all assessment results	0.011	0.009	0.009	0.783	0.011	0.010	0.009	0.786

Table shows means of selected baseline variables for each treatment group in the baselined sample (columns 1-3) and endlined sample (columns 5-7). Column 4 shows p-values for equal means in the baselined sample, evaluating balanced treatment assignments. Column 8 shows p-values for equal means in the endlined sample, evaluating balanced attrition. See footnote to Table D.1 for details on variable definitions. Hypothesis tests are based on heteroskedasticity-robust standard errors clustered by treatment date.

Table D.4: Summary Statistics for Experimental and External Comparison Samples

	QLFS SA	QLFS Johannesburg			Experimental
		All	Age-restricted	Reweighted	Sample
Age	36.5	37.4	26.5	23.6	23.7
	(12.7)	(11.9)	(4.7)	(3.3)	(3.3)
Male	0.492	0.513	0.500	0.381	0.382
Black	0.796	0.786	0.824	0.983	0.983
Highest Education Level					
Less than Secondary	0.567	0.430	0.388	0.011	0.011
Completed Secondary	0.296	0.362	0.432	0.610	0.610
More than Secondary	0.127	0.188	0.163	0.378	0.379
Employed	0.468	0.566	0.445	0.373	0.378
Searching	0.319	0.519	0.536	0.532	0.968
Earnings	971	1379	888	709	187
	(12766)	(10871)	(3158)	(2300)	(501)

Table compares the sample of workseekers in this study (column 5) to several external benchmarks: the country (column 1), the metro area of Johannesburg where the study takes place (column 2), people in Johannesburg in the eligible age range for the study (column 3), and people in Johannesburg in the eligible age range for the study, reweighted with propensity scores to approximate the experimental sample on age, education, sex, and race (column 4). National and metro area statistics are calculated from the Quarterly Labour Force Survey (QLFS), averaging over all 2016 and 2017 waves and using post-stratification weights provided by Statistics South Africa. The external benchmarks in columns 1 and 2 use only people aged 18-65 to approximate the working-age population. Standard deviations are shown in parentheses for all continuous variables. Earnings are for the last week and are in South Africa Rands. 1 Rand $\approx$ USD 0.167 in purchasing power parity terms.

D.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 D.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 poverty lines based on 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 rand or USD 232 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 7-9% of the monthly minimum wage at the time of the study.

Table D.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	Early 2016	1386	232	320	54
Household (4 people)	Early 2016	5544	927	1279	214
Minimum wage					
Domestic work	2015-2016	2550	427	588	98
Hospitality	2015-2016	2750	460	634	106
Wholesale and retail	2015-2016	3250	544	750	125
Private security/contract cleaning	2015-2016	3500	585	808	135

Panel B: Benchmarking sample earnings and certification treatment effects on earnings

		Weekly		As % of poverty line		As % of min. wage	
	Date	ZAR	USD	Adult	Household	Hospitality	Retail
Baseline mean earnings if employed	Late 2016	562	94.1	1.76	0.44	0.89	0.75
Endline mean earnings	Early 2017	159	26.6	0.50	0.12	0.25	0.21
Endline mean earnings if employed	Early 2017	518	86.7	1.62	0.41	0.82	0.69
Treatment effect	Early 2017	54.1	9.05	0.17	0.04	0.09	0.07

Calculations assume 1 rand ≈ 0.167 USD in purchasing power parity terms; 4.33 weeks per month. Household poverty lines assume households of four people with only one earner. 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) and minimum wages are from the Department of Labor for 2015. Minimum wages are for large urban areas (Area A). They are for hospitality businesses with less than 10 employees and shop assistants in the wholesale and retail sector.

D.3 Non-response

The phone survey after 3-4 months is our main source of endline data. We use a 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 D.6). Non-response does not differ over most baseline characteristics (Table D.7). 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 D.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. Heteroskedasticity-robust standard errors clustered by treatment date are shown in parentheses.

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

	(1) Text Message Survey	(2) Endline Phone Survey
Completed secondary education only	-0.010	-0.004
	(0.013)	(0.005)
Numeracy score	-0.031	0.002
	(0.006)	(0.003)
Communication score	0.008	0.005
	(0.005)	(0.003)
Concept formation score	-0.020	0.001
	(0.006)	(0.003)
Grit score	-0.002	-0.007
	(0.005)	(0.003)
Other scores	0.002	-0.001
	(0.004)	(0.003)
Perceived numeracy score	-0.000	-0.000
	(0.000)	(0.000)
Perceived literacy score	0.012	-0.003
	(0.010)	(0.005)
Perceived concept formation score	0.006	-0.004
	(0.009)	(0.005)
Self-esteem index	0.003	0.002
	(0.005)	(0.002)
Age	-0.002	0.001
	(0.002)	(0.001)
Male	0.052	0.014
	(0.011)	(0.006)
Employed	-0.008	-0.003
	(0.009)	(0.005)
Above median discount factor	0.009	0.005
	(0.009)	(0.005)
Individual is present biased	0.015	0.008
	(0.011)	(0.006)
Above median risk aversion	0.000	0.000
	(0.009)	(0.006)
p: All coefficients jointly zero	0.000	0.109
Mean outcome		
# observations	5985	5985
# clusters	82	82

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. Heteroskedasticity-robust standard errors clustered by treatment date are shown in parentheses.

D.4 Additional Treatment Effects

Table D.8 shows the public certification effects of our main outcomes without conditioning on the prespecified covariates. Table D.9 shows the public certification effects on the same outcomes conditional on the two covariates that are unbalanced at baseline: search and earnings. The results are very similar across all sets of covariates.

Table D.10 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, which expand on the discussion in footnote 22 of the paper. 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, although the different survey methods mean the time comparison should be interpreted cautiously. 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 months 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 average self-esteem at either point in time or the distribution of self-esteem at endline (Figure D.1).

The difference in results between the two surveys is not driven by differences in sample selection. To show this, we estimate treatment effects on beliefs in the endline phone survey using the sample of workseekers who responded to the text message survey. The results are almost identical to those using the sample of workseekers who responded to the endline phone survey (columns 1 and 2 versus 4 and 5 of Table D.10).

The measures in Table D.10 capture candidates' beliefs about their performance on the assessments they took. These do not necessarily match their beliefs about their skills. For example, a candidate may believe that they have good numeracy skills but performed poorly in the numeracy assessment as they were very tired that day. If beliefs about assessment results and beliefs about skills are weakly correlated, then our belief measures may not capture workseekers' decision-relevant beliefs. To address this possibility, we ask candidates if their communication and numeracy skills are in the top, middle, or bottom third of people aged 18-34, from disadvantaged backgrounds, with high school education (the population typically assessed by Harambee). This is not a question about their result on a specific assessment. Treatment increases the share of the two skills where candidates' beliefs about their domain-specific skills match their actual assessment results by 12.4 percentage points (standard error 2.2 p.p.). This is only slightly lower than the treatment effect on the share of the skills were candidates' beliefs about their assessment results match their actual

Table D.8: Treatment Effects on Key Outcomes Without Covariates

	(1) Employed	(2) Earnings[c]	(3) Skill belief accurate	(4) Targeted search	(5) Used report[b]
Public treatment	0.046	0.336	0.155	0.045	0.699
	(0.013)	(0.076)	(0.010)	(0.010)	(0.013)
Private treatment	0.001	0.147	0.117	0.046	0.288
	(0.014)	(0.078)	(0.010)	(0.012)	(0.012)
Mean outcome	0.309	159.291	0.389	0.155	0.000
Mean outcome for employed		518.291			
# observations	6607	6589	6607	6609	6609
# 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 outcomes are for the control group. All outcomes use a 7-day recall period. 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.

assessment results (15.8 percentage points with standard error 0.8 percentage points). This shows that candidates update beliefs about their skills more generally, not just updating beliefs about their performance on the assessments. Because this is not a primary outcome, we collect this measure only for a random 50% sample of the first 3,000 candidates to complete the survey. We ask only about communication and numeracy because we expect candidates to have the most precise beliefs about these prominent skills.

Table D.11 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). We discuss these treatment effects in Sections 5.2 and 5.3 of the paper.

Table D.12 reports public and private certification effects on all prespecified 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 is 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. None of the treatment effects on these averages provides substantively different information to the treatment effects on individual outcomes.

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 D.9: Treatment Effects on Key Outcomes With Additional Covariates

	(1) Employed	(2) Earnings[c]	(3) Skill belief accurate	(4) Targeted search	(5) Used report[b]
Public treatment	0.053	0.348	0.158	0.051	0.699
	(0.012)	(0.074)	(0.008)	(0.010)	(0.013)
Private treatment	0.011	0.160	0.124	0.047	0.290
	(0.012)	(0.076)	(0.008)	(0.010)	(0.013)
Mean outcome	0.309	159.291	0.389	0.155	0.000
Mean outcome for employed		518.291			
# observations	6607	6589	6607	6609	6609
# clusters	84	84	84	84	84

Coefficients are from regressing each outcome on a vector of treatment assignments and randomization block fixed effects, prespecified covariates, and two covariates that are unbalanced at baseline but not prespecified (search and earnings). Heteroskedasticity-robust standard errors shown in parentheses, clustering by treatment date. Mean outcomes are for the control group. All outcomes use a 7-day recall period. 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 D.13 shows the decomposition of both public and private certification effects into extensive and intensive margin effects. Figure D.2 shows the quantile treatment effects of public and private certification on earnings. The table and figure allow comparison of the private and public effects on labor market outcomes at different margins.

Table D.14 shows the distribution of earnings conditional on employment in each treatment group, with and without reweighting to adjust for differences across groups in selection into employment. This table shows that earnings conditional on employment are slightly higher in the private than public certification group.

Table D.10: Treatment Effects on Self-Beliefs through Time

	Perceived numeracy tercile correct			Above-median self-esteem		
	(1)	(2)	(3)	(4)	(5)	(6)
Public	0.233	0.233	0.316	0.002	-0.002	-0.001
	(0.013)	(0.015)	(0.015)	(0.013)	(0.015)	(0.015)
Private treatment	0.200	0.205	0.333	-0.002	0.001	0.017
	(0.015)	(0.016)	(0.016)	(0.015)	(0.018)	(0.015)
p: public = private	0.010	0.043	0.251	0.812	0.859	0.238
Mean outcome	0.396	0.404	0.399	0.553	0.558	0.479
# observations	6601	5292	5297	6609	5027	5027
# clusters	84	84	84	84	84	84
Survey round	Phone	Phone	Text	Phone	Phone	Text
Sample from survey round	Phone	Text	Text	Phone	Text	Text

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 outcomes are 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 (4) report results from the main phone follow-up survey. Columns (3) and (6) report results from the text message survey conducted 2-3 days after treatment. Columns (2) and (5) report results from main phone follow-up survey for the subsample respondents who answered both surveys. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

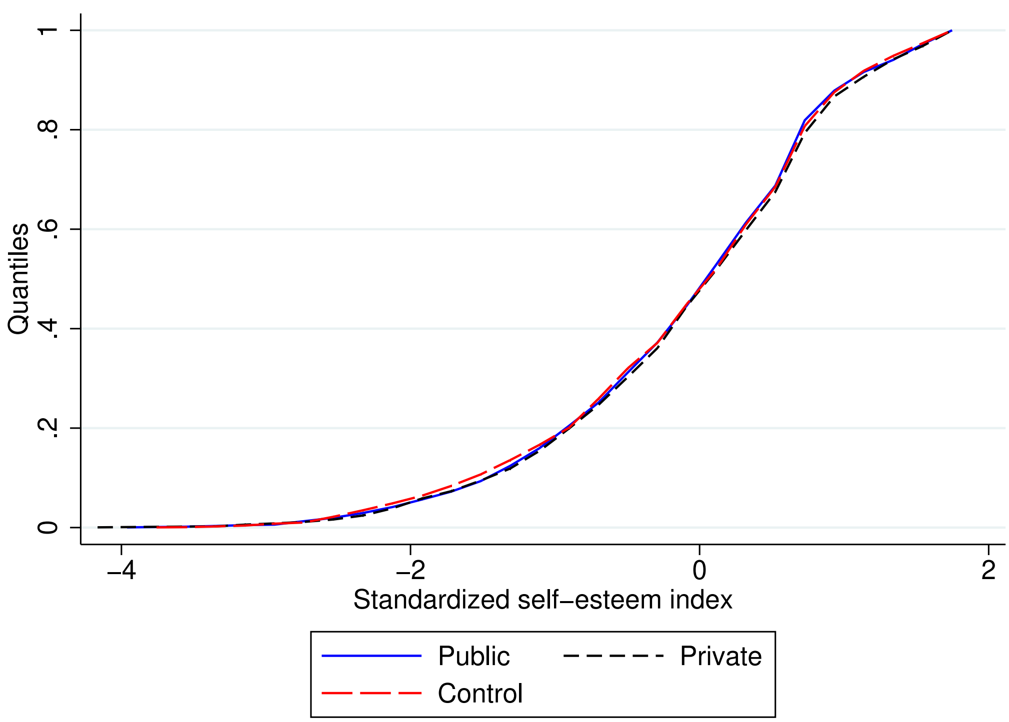


Figure D.1: Distribution of Self-Esteem at Endline by Treatment Group

Table D.11: Heterogeneous Treatment Effects on Employment

	(1)	(2)	(3)
Panel A: Heterogeneous Treatment Effects by Single Index Skill Measures			
Public treatment	0.052	0.052	0.053
	(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)
Mean outcome	0.309	0.309	0.309
# observations	6607	6607	6603
# clusters	84	84	84
Panel B: Heterogeneous Treatment Effects by Alternative Information Sources			
Public treatment	0.051	0.052	0.051
	(0.011)	(0.012)	(0.012)
× post-secondary education	-0.028		
	(0.028)		
× employed at baseline		-0.043	
		(0.032)	
× $\hat{\Pr}$(Employed at endline $\|$X)			-0.076
			(0.028)
Mean outcome	0.309	0.309	0.309
# 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 in panels A and B 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 D.12: 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	-0.013	-0.020	0.019	-0.036	-0.094
	(0.032)	(0.014)	(0.042)	(0.048)	(0.080)
Private treatment	0.006	-0.006	0.037	-0.036	-0.033
	(0.032)	(0.014)	(0.038)	(0.049)	(0.088)
q: Public effect = 0	0.530	0.972	0.972	0.972	0.972
q: Private effect = 0	0.749	1.000	1.000	1.000	1.000
q: Public = private effect	0.849	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 treatment	0.019	0.016	0.013	-0.005	0.001
	(0.026)	(0.022)	(0.013)	(0.004)	(0.004)
q: Public effect = 0	0.530	1.000	1.000	1.000	1.000
q: Private effect = 0	0.463	1.000	1.000	1.000	1.000
q: Public = private effect	0.864	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) Used report[b]	(2) Applications with report[b,c]	(3) Interviews with report[b,c]	(4) Offers with report[b,c]
Public	0.699	1.682	0.432	0.112
	(0.013)	(0.040)	(0.023)	(0.011)
Private treatment	0.290	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 treatment	0.050	0.011	0.029	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	0.001
q: Private effect = 0	0.138	0.509	0.132	0.509	0.339
q: Public = private effect	0.002	0.003	0.133	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	0.337	0.197	0.020
	(0.028)	(0.074)	(0.039)	(0.010)
Private treatment	0.069	0.162	0.094	0.017
	(0.030)	(0.078)	(0.046)	(0.009)
q: Public effect = 0	0.001	0.001	0.001	0.019
q: Private effect = 0	0.103	0.068	0.068	0.068
q: Public = private effect	0.525	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 q-values in the first column of each panel adjust for multiple testing across the four indices. 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 outcomes are 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.

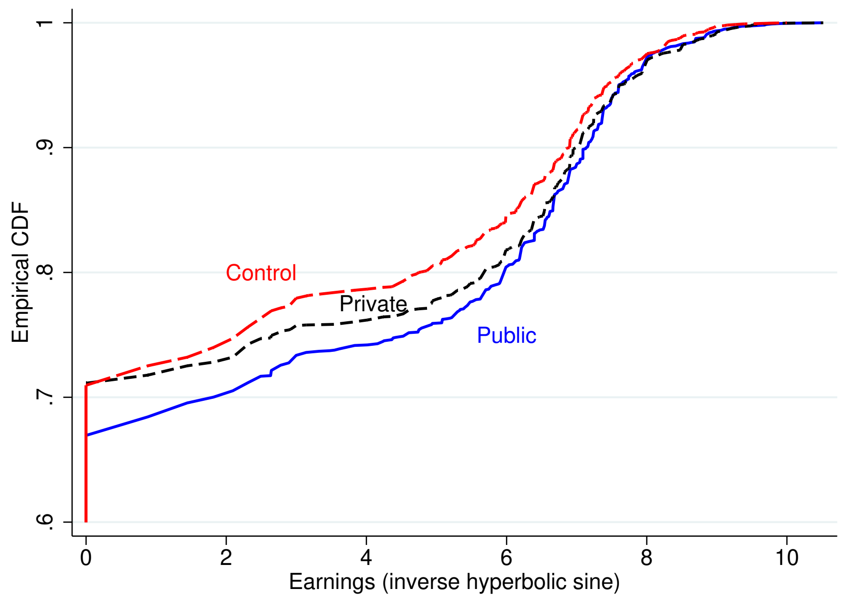
Table D.13: Treatment Effects on Labor Market Outcomes at Extensive and Intensive Margins

	(1) Hours[c]	(2) Earnings[c]	(3) Hourly wage[c]	(4) Written contract
Panel A: Public Treatment Effects				
Total effect	0.201	0.337	0.197	0.020
	(0.052)	(0.073)	(0.039)	(0.010)
Extensive margin	0.188	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)	(0.027)	(0.008)
Treatment effect conditional	0.037	0.194	0.158	-0.001
on employment	(0.058)	(0.113)	(0.078)	(0.024)
Panel B: Private Treatment Effects				
Total effect	0.066	0.162	0.094	0.017
	(0.047)	(0.077)	(0.046)	(0.009)
Extensive margin	0.041	0.058	0.030	0.004
	(0.043)	(0.062)	(0.033)	(0.005)
Intensive margin	0.025	0.103	0.064	0.013
	(0.019)	(0.039)	(0.029)	(0.007)
Treatment effect conditional	0.083	0.339	0.209	0.041
on employment	(0.063)	(0.128)	(0.095)	(0.024)
Panel C: Testing Equality of Public & Private Effects				
Total effect	0.009	0.025	0.026	0.768
Extensive margin	0.001	0.001	0.001	0.001
Intensive margin	0.529	0.380	0.791	0.102
Treatment effect |employment	0.440	0.234	0.585	0.078

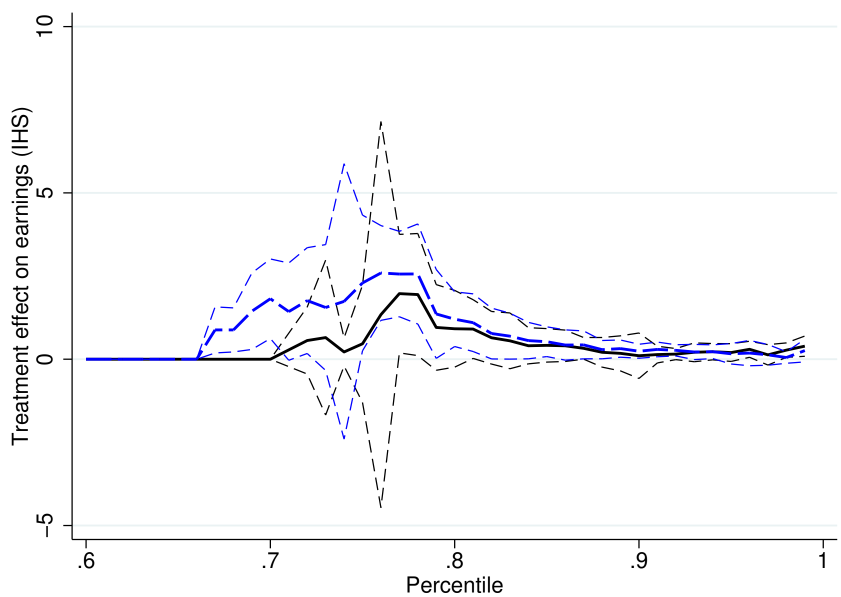
This table reports decompositions of public and private treatment effects on job characteristics into extensive and intensive margin effects. The extensive margin effects are the treatment effects on job characteristics due to the treatment effect on employment, evaluated at the mean job characteristics for the control group. The intensive margin effects are the differences between the treatment effects and extensive margin effects, which must be due to changes in job characteristics for the employed candidates in the treatment group. The conditional effect is the implied mean change in job characteristics per employed treatment group candidate. Heteroskedasticity-robust standard errors are shown in parentheses, clustering by treatment date. All outcomes use a 7-day recall period. Outcomes marked with [c] use the inverse hyperbolic sine transformation.

Figure D.2: Quantile Treatment Effects on Earnings

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



Panel B: Quantile Treatment Effects of Public and Private Certification on Earnings



Panel A shows the empirical distributions of earnings in the control, private certification, and public certification groups. Earnings are the inverse hyperbolic sine transformation of earnings in South African rand, with 1 rand $\approx$ 0.167 USD 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. Panel B shows the quantile treatment effects (QTEs) of public and private 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.

Table D.14: Earnings Distributions by Treatment Group Adjusting for Observed Covariates

Sample	Probability of employment	Earning distribution for employed			
		Mean	Std dev.	25^{th} pctile	75^{th} pctile
Control group	0.307	5.177	2.547	2.776	7.090
Private group	0.302	5.753	2.379	4.931	7.244
Private group reweighted	0.302	5.804	2.333	4.942	7.244
Public group	0.348	5.458	2.520	3.577	7.244
Public group reweighted	0.348	5.515	2.520	3.832	7.090

This table shows the distribution of earnings conditional on employment for each treatment group. The rows marked show the earning distribution in group X after reweighting the group to have the same distribution of baseline covariates (measured skills, self-reported skills, education, age, gender, employment, discount rate, risk aversion).

E Audit Study

We conduct an audit study to identify the effect of information provision on firm decisions, without any scope for mediating behavior by workseekers. We submit real workseekers' applications to entry-level job vacancies and randomly vary the information firms see about workseekers' skills. This appendix reports more information about the process and sample to help interpret the results reported in Section 4.3.

We implement the audit study in nine sequential rounds. In each round, we invite candidates by text message to submit application materials to us, within 7 days, for an undisclosed job opportunity.[37] We do not explicitly indicate our affiliations or link the message to Harambee. We send one reminder text message to all candidates 1-3 days after the initial invitation.

We invited 2,220 candidates to send CVs over the nine rounds. We randomly sample candidates from those who had already completed the workseeker survey. 717 candidates (28%) submit CVs within the one week period. Most CVs include some information about proxies for candidates' skills: 91% include a reference letter or contact information for referees and 55% include their secondary school graduation results (Table E.1, panel A). The 717 responders are similar to the full workseeker sample on all baseline covariates except gender, where deliberately oversampled men for an even gender split. Candidates in the private treatment group are slightly more likely to respond to the invitation (Panel B). All treatment effects are robust to reweighting the responders to have the same distribution of treatment assignments and baseline covariates as the full workseeker sample.

For each application received, we record information on when the application was received, where it was sent from, what documents are included, and an indicator for scan quality of included documents (e.g. photographs versus high-quality scans). We also send the candidate an acknowledgement of receipt.

Simultaneously, we compile job vacancies from several online job posting sites. We selected only vacancies suitable for entry-level workers, so that all candidates in our sample are eligible to apply. We exclude jobs that look suspicious or are discriminatory, for example: jobs that ask for payments of any kind to apply, promise unrealistic salaries or benefits, or discriminate based on appearance, race, or gender. This generates a sample of 1,068 vacancies over the nine rounds, though we exclude 70 vacancies for reasons discussed below. Among the vacancies, 48% are for sales jobs, with the remaining vacancies spread over clerical, call center, factory, restaurant and retail jobs.

We submit 4 applications to each vacancy, each "from" a different candidate using a different email address. We do not represent ourselves as the candidate. Instead, we use a generic email address designed to look like the application was scanned at a copy/printing shop, a generic subject

[37] 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 in South Africa is generally understood to include all materials relevant to job applications.

Table E.1: Comparison Between Audit and Workseekers Study Samples

	Workseekers in audit sample			All workseekers		
	Mean	Std Dev.	Obs	Mean	Std Dev.	Obs
Panel A: Characteristics of responses received from workseekers						
Includes references or a reference letter	0.91	0.29	713	-	-	-
Includes a copy of ID document	0.47	0.50	714	-	-	-
Includes information about secondary school completion	0.55	0.50	714	-	-	-
Panel B: Characteristics of workseekers						
Public treatment	0.30	0.46	717	0.33	0.47	6891
Private treatment	0.37	0.48	717	0.31	0.46	6891
Age	23.2	3.12	717	23.6	3.30	6891
Male	0.48	0.50	717	0.38	0.49	6891
University degree / diploma	0.18	0.38	717	0.17	0.37	6891
Any other post-secondary qualification	0.24	0.42	717	0.21	0.41	6891
Completed secondary education only	0.58	0.49	717	0.61	0.49	6891
Numeracy assessment score (z-score)	0.06	0.96	717	0.00	1.00	6891
Literacy/communications assessment score (z-score)	0.02	0.94	717	0.00	1.00	6891
Concept formation assessment score (z-score)	0.11	0.93	717	0.00	1.00	6891
Grit assessment score (z-score)	0.10	0.99	717	0.00	1.00	6891
Worked in the last 7 days (endline)	0.40	0.49	717	0.38	0.48	6891

line, and generic email message.[38] We send most applications within 2 weeks of compiling the vacancy list.

We use a three-stage randomization process. First, we generate multiple applications per candidate and randomly assign half of these to treatment status and half to control status. Treatment applications are sent with a public certificate and control applications without any certificate. In all other respects, treatment and control applications are identical. This randomization is independent of workseekers' treatment status in the workseekers' study. This generates within- and between-candidate variation in the information content of their applications. Second, we randomize vacancies to receive either one or three applications with certificates. This generates within-vacancy variation in the information content of the applications received and between-vacancy variation in the overall information environment. Third, we randomly match applications to vacancies, subject to the target number of treated and control applications and the constraint that no candidate's application is sent to the same vacancy more than once. The realized distribution of treatment assignments shown in Table E.2, Panel A matches the intended design: half of the applications are sent with certificates and, mechanically, applications sent with certificates are three times more likely to be sent to vacancies that receive three applications with certificates.

We monitor and record responses for two weeks after sending the applications. We classify each response into one of these categories: (1) interview invitation, (2) request to send more information

[38] We cross-randomize the subject lines "Application for <vacancy>" and "Application for <candidate name>" with the email messages "Please find attached the application for <vacancy> as recently advertised online" and "Please find the application for <candidate name> for <vacancy>, as recently advertised online."

Table E.2: 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	3992
Had one report in a vacancy with three reports	0.38	0.48	3992
Had no report in a vacancy with one report	0.37	0.48	3992
Had no report in a vacancy with three reports	0.13	0.33	3992
Panel B: Responses to applications submitted			
Any response received	0.15	0.35	3992
Interview request received	0.09	0.29	3992

or visit the establishment in person, (3) email bounce, (4) scam, and (5) other - mostly personalized acknowledgements of receipt. If any application sent to a vacancy receives a type (3) or (4) response, we drop the vacancy from the sample. We define two outcome variables for analysis. First, any application that receives a type (1) response is coded as an 'interview invitation.' Second, any response that receives a type (1), (2), or (5) response is classified as 'any response'. We forward all responses to the relevant candidate so they can contact the firm. We do not monitor the outcome of the candidate-firm interaction after this point, because interview invitations are too rare to allow us to precisely estimate treatment effects on post-interview outcomes.

The final sample consists of 3,992 applications sent to 998 vacancies, after dropping 70 vacancies with bounce or scam responses. Of these applications, 15% receive any response, including 9% that receive interview invitations (Table E.2, panel B).

F Placebo Certification Experiment: Sample Certificate and Treatment Effects

Figure F.1: Sample Placebo Certificate

THE WORLD BANK

harambee
YOUTH EMPLOYMENT ACCELERATOR
WORK FOR WORK

REPORT ON ASSESSMENT PROCESS

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 «date».

«**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.

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. 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.

This figure shows an example of the certificates given to candidates in the placebo 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, an email certificate, and guidelines on how to request more certificates.

Table F.1: Public and Placebo Certification Effects

	(1) Labor market index	(2) Employed	(3) Hours^{c}	(4) Earnings^{c}	(5) Hourly wage^{c}	(6) Written contract
Public	0.120	0.052	0.201	0.337	0.197	0.020
	(0.027)	(0.012)	(0.052)	(0.074)	(0.039)	(0.010)
Placebo	0.027	0.020	0.040	0.068	0.053	0.005
	(0.043)	(0.028)	(0.075)	(0.185)	(0.129)	(0.021)
p: public = placebo	0.041	0.245	0.045	0.147	0.267	0.472
Placebo / public ratio	0.221	0.376	0.197	0.202	0.271	0.240
# 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 outcomes are for the control group. All outcomes use a 7-day recall period. 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 0.257 for the 5 labor market outcomes. The sample sizes differ across columns due to item non-response, mostly from respondents reporting that they don't know the answer.

Table G.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		
Uses external recruiting services	69	1.75	0.43	1.00	2.00
# 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
Mean monthly compensation for employees in last financial year	58	8,447	16273	2,500	9,000
Total payroll costs in last financial year (millions)	31	1.28	2.77	0.08	3.20

Table shows summary statistics for selected firm attributes variables. Percentiles are omitted for binary variables. First six rows are indicators for sectors. All monetary figures are reported in South Africa Rands. 1 Rand ≈ USD 0.167 in purchasing power parity terms. # observations varies due to item non-response. Missing values for the final 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.

G 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 (WTP) 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 G.1 reports summary statistics for this sample.

We measure WTP using a standard Becker-DeGroot-Marschak mechanism. We first explain the entire mechanism, then run a practice round with a bar of chocolate, and then run the mechanism for the database.

For the database round of WTP, we first describe the database and show them a live demonstration. Figures G.1 and G.2 shows screenshots of the platform marketed to firms. Second, we explain the mechanism and ask respondents their WTP. Third, we tell them the 'normal' price of 10,000 South African Rands (USD 1,670 PPP) for three months access. Fourth, we ask if they want to revise their initial WTP after learning the 'normal' price. The 'normal' price we state is not a market-determined price, as this was a new product we were piloting with Harambee. If their updated WTP is higher than the normal price minus the discount, we give them access to the database. If their updated 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.

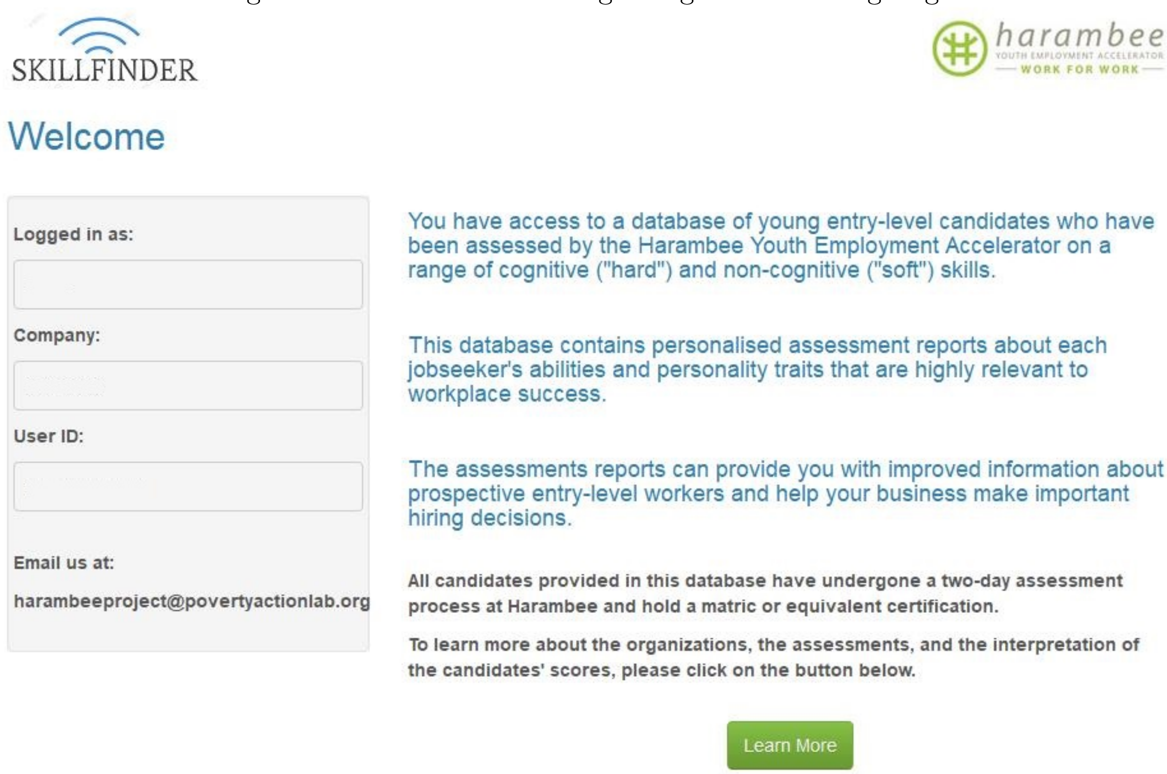
Figure G.3 shows the distribution of updated WTP. The distribution is similar using initial WTP. After learning the price, only 22 and 5% of respondents updated their WTP respectively upward and downward. All the downward revisions were respondents whose initial WTP was above the 'normal' price we quoted. The share of firms with positive WTP is the same before and after updating. The mean WTP is 670 South African Rands (USD 112 PPP) higher before updating, due to one large downward revision by a firm whose initial WTP was five times higher than the 'normal' price.

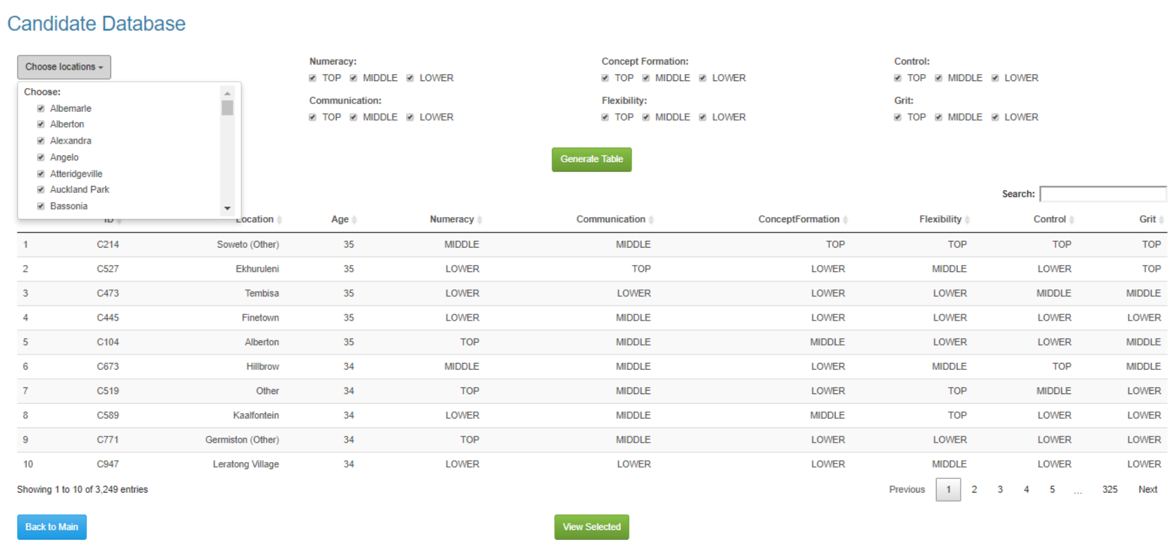
WTP is robustly higher for firms who plan to hire an entry-level worker in the next year. It is not robustly associated with any other firm characteristic listed in Table G.1, using either OLS or LASSO analyses.

To elicit these firms' preferences for different types of skills, 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, in line with 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 G.2). All six profiles' median rank is between second and fourth. The share of firms ranking each profile highest ranges from 6 to 33%. The seventh profile has middle terciles for all six assessments and has a one-year post-secondary education 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.

We conduct a second experiment where we ask firms to rank profiles with assessment results shown for some skills and concealed for others. This assesses 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.

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





	ID	Location	Age	Numeracy	Communication	ConceptFormation	Flexibility	Control	Grit
1	C214	Soweto (Other)	35	MIDDLE	MIDDLE	TOP	TOP	TOP	TOP
2	C527	Ekhuruleni	35	LOWER	TOP	LOWER	MIDDLE	LOWER	TOP
3	C473	Tembisa	35	LOWER	LOWER	LOWER	LOWER	MIDDLE	MIDDLE
4	C445	Finetown	35	LOWER	MIDDLE	LOWER	LOWER	LOWER	LOWER
5	C104	Alberton	35	TOP	MIDDLE	MIDDLE	LOWER	LOWER	MIDDLE
6	C673	Hillbrow	34	MIDDLE	MIDDLE	LOWER	MIDDLE	TOP	MIDDLE
7	C519	Other	34	TOP	MIDDLE	LOWER	TOP	MIDDLE	LOWER
8	C589	Kaalfontein	34	LOWER	MIDDLE	MIDDLE	TOP	LOWER	LOWER
9	C771	Germiston (Other)	34	TOP	MIDDLE	LOWER	LOWER	LOWER	LOWER
10	C947	Leratong Village	34	LOWER	LOWER	LOWER	MIDDLE	LOWER	LOWER

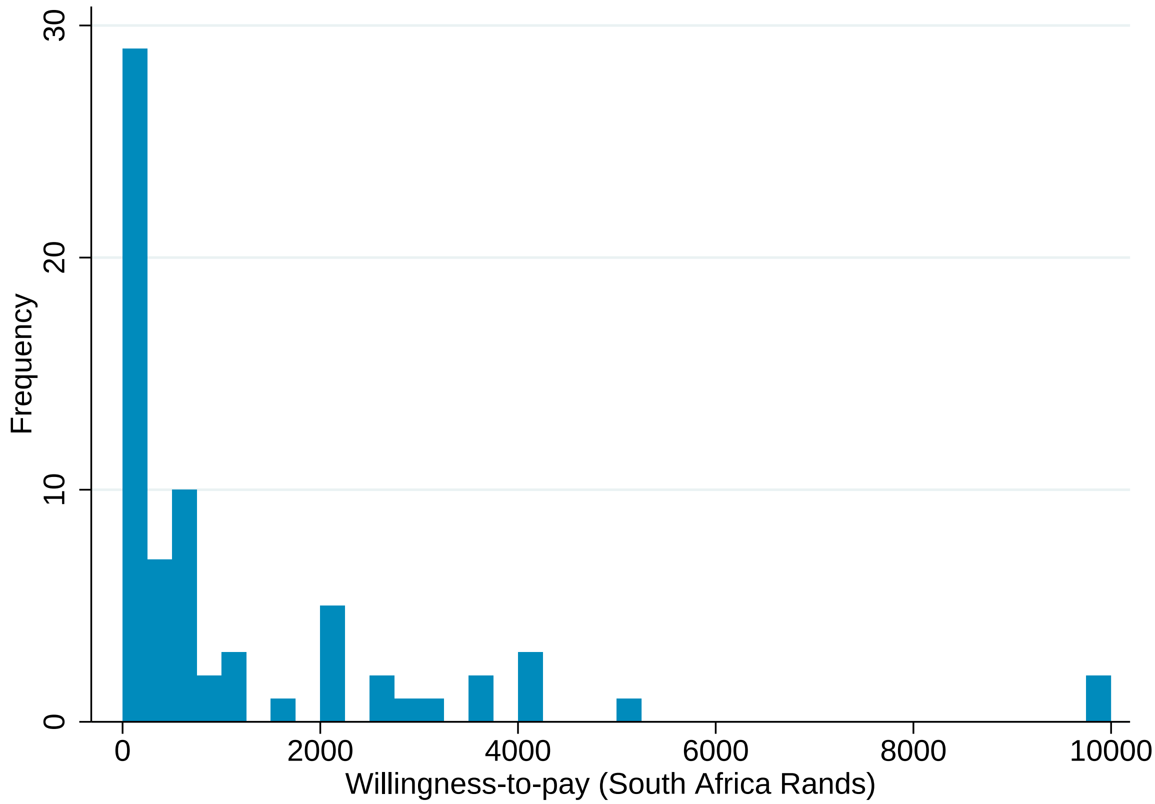
Figure G.2: Screenshot of Individual Candidate Profile on Platform

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

Profile content		(1)	(2)	(3)
		Share of firms ranking profile		Median
Top tercile	Highest education	First	Last	ranking
Communication	Completed secondary school	0.119	0.015	3
Concept formation	Completed secondary school	0.075	0.030	4
Focus	Completed secondary school	0.328	0.060	3
Grit	Completed secondary school	0.134	0.045	4
Numeracy	Completed secondary school	0.060	0.090	2
Planning	Completed 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 assessment results and different levels of education. All profiles have middle terciles for skills except that listed in the first column.

Figure G.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 G.1 and G.2. Values are in South African rand, with 1 rand ≈ USD 0.167 in purchasing power parity terms. The maximum possible bid is 10,000 South African rand.

Appendix References

ABEBE, G., S. CARIA, AND E. ORTIZ-OSPINA (2020): "The Selection of Talent: Experimental and Structural Evidence from Ethiopia," Manuscript, University of Bristol.

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.

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.

BEAMAN, L., N. KELEHER, AND J. MAGRUDER (2018): "Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi," *Journal of Labor Economics*, 36, 121–153.

BENJAMINI, Y., A. KRIEGER, AND D. YEKUTIELI (2006): "Adaptive Linear Step-Up Procedures That Control the False Discovery Rate," *Biometrika*, 93, 491–507.

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.

CASALE, D. AND D. POSEL (2011): "English Language Proficiency and Earnings in a Developing Country: The Case of South Africa," *Journal of Socio-Economics*, 40, 385–393.

CHAMORRO-PREMUZIC, T. AND A. FURNHAM (2010): *The Psychology of Personnel Selection*, Cambridge University Press.

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.

DIAMOND, A. (2013): "Executive Functions," *Annual Review of Psychology*, 64, 135–168.

DU RAND, G., H. VAN BROEKHUIZEN, AND D. VON FINTEL (2011): "Numeric Competence, Confidence and School Quality in the South African Wage Function," Manuscript, Stellenbosch 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.

Ederer, P., L. Nedelkoska, A. Patt, and S. Castellazzi (2015): "What Do Employers Pay for Employees' Complex Problem Solving Skills?" *International Journal of Lifelong Education*, 34, 430–447.

Eskreis-Winkler, L., A. 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.

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

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

Hanushek, E., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015): "Returns to Skills Around the World: Evidence from PIAAC," *European Economic Review*, 73, 103–130.

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

Heckman, J. and T. Kautz (2012): "Hard Evidence on Soft Skills," *Labour Economics*, 19, 451–464.

Heckman, J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Non-Cognitive Abilities on Labour Market Outcomes and Social Behaviour," *Journal of Labour Economics*, 24, 411–482.

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

Heppner, P. (1988): *The Problem Solving Inventory (PSI): Manual*, Palo Alto: Consulting Psychologists.

Heppner, P., T. Pretorius, M. Wei, D. Lee, and Y. Wang (2002): "Examining the Generalizability of Problem-Solving Appraisal in Black South Africans," *Journal of Counseling Psychology*, 49, 484–498.

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

Kalechstein, A., T. Newton, and W. van Gorp (2003): "Neurocognitive Functioning is Associated with Employment Status: A Quantitative Review," *Journal of Clinical and Experimental Neuropsychology*, 25, 1186–1191.

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.

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.

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

Nisbett, R. (2009): *Intelligence and How to Get It: Why Schools and Cultures Count*, New York: W. W. Norton and Company.

Posner, M. and G. DiGirolamo (1998): "Executive Attention: Conflict, Target Detection, and Cognitive Control," in *The Attentive Brain*, ed. by R. Parasuraman, MIT Press, 401–423.

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

Pretorius, T. (1993): "Assessing the Problem-Solving Appraisal of Black South African Students," *International Journal of Psychology*, 28, 861–870.

Rath, J., D. Langenbahn, D. Simon, R. L. Sherr, J. Fletcher, and L. Diller (2004): "The Construct of Problem Solving in Higher Level Neuropsychological Assessment and Rehabilitation," *Archives of Clinical Neuropsychology*, 19, 613–635.

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

Roberts, R., G. Goff, F. Anjoul, P. Kyllonen, G. Pallier, and L. Stankov (2000): "Armed Services Vocational Aptitude Battery (ASVAB): Little More than Acculturated Learning?" *Learning and Individual Differences*, 12, 81–103.

Schmidt, F. L. and J. E. Hunter (1998): "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings," *Psychological Bulletin*, 124, 135–168.

Schmidt, F. L., I.-S. Oh, and J. Shaffer (2016): "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research Findings," Manuscript, University of Iowa.

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

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

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.

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.