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 Communication	Assessment result for Concept formation	Assessment result for Numeracy	Career aptitude profile	Criminal record check	High school graduation certificate	Reference
Hospitality	11	0.82	1.00	0.91	0.64	0.91	0.64	0.00
Retail	16	0.69	0.56	0.88	0.81	0.94	0.75	0.06
Corporate	6	1.00	1.00	0.83	0.33	1.00	0.00	0.00
Total	33	0.79	0.91	0.79	0.67	0.94	0.58	0.03

Table shows use of assessment results and other information by 33 firms that have long-term recruiting relationships with Harambee. Firms 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

Panel A: Correlations In First 17 Days of Assessment (1615 workseekers)					
	Concept formation	Grit	Numeracy	Control	Flexibility
Communication	0.337	0.127	0.386	0.237	0.126
Concept formation		0.108	0.489	0.174	0.098
Grit			0.162	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:

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

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

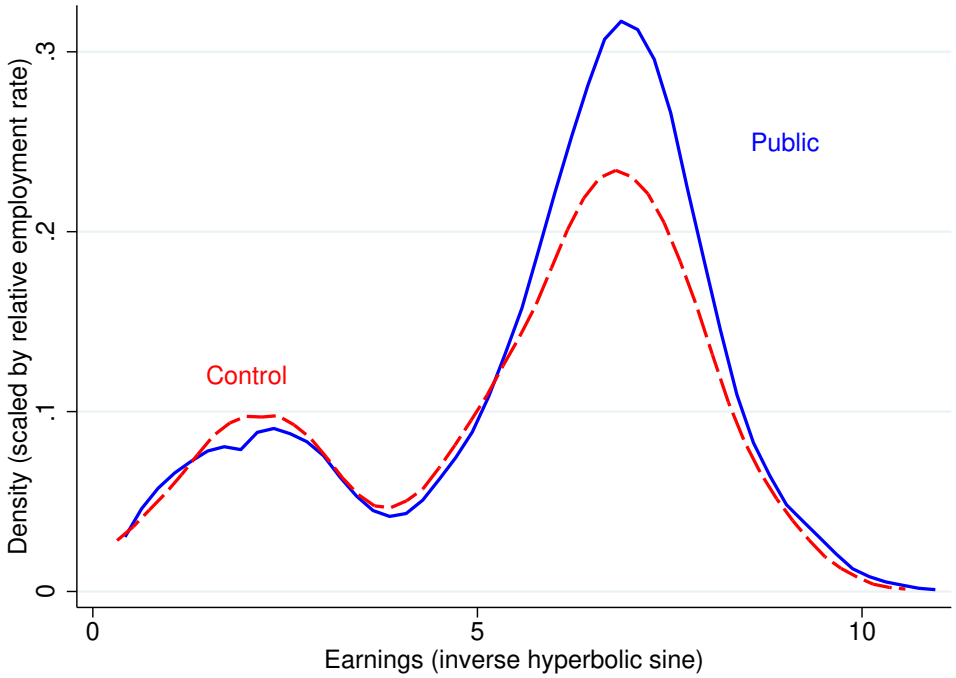
[36] Attanasio et al. (2011) show that the intensive margin effect can be further decomposed into two terms: the treatment effect on earnings conditional on candidates' baseline characteristics, and the difference in baseline characteristics between employed candidates in the treatment and control groups. 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 $\approx$ 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 $\approx$ 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 $\approx$ 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)	(2)
	Text Message Survey	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.

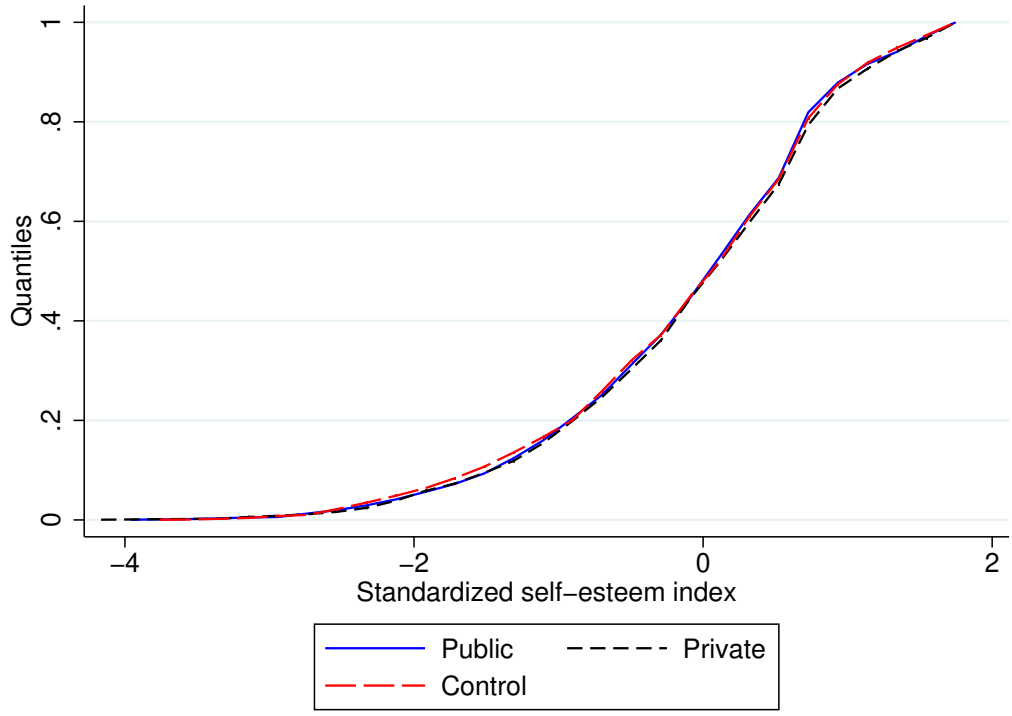


Figure D.1: Distribution of Self-Esteem at Endline by Treatment Group

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.

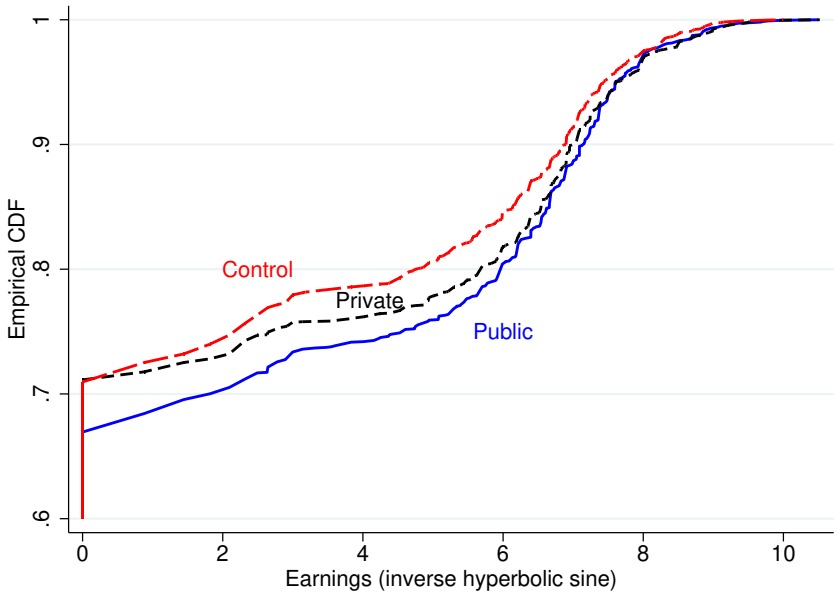
Table D.13: Treatment Effects on Labor Market Outcomes at Extensive and Intensive Margins

	(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 on employment	0.037	0.194	0.158	-0.001
	(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 on employment	0.083	0.339	0.209	0.041
	(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

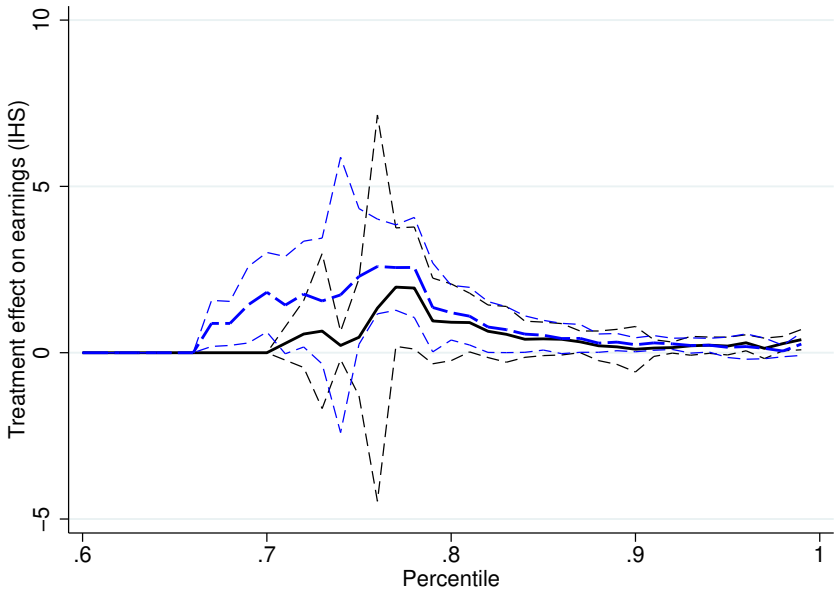
This table reports decompositions of public and private treatment effects on job characteristics into extensive and intensive margin effects. The extensive margin effects are the treatment effects on job characteristics due to the treatment effect on employment, evaluated at the mean job characteristics 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^{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

	Probability of	Earning distribution for employed			
Sample	employment	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

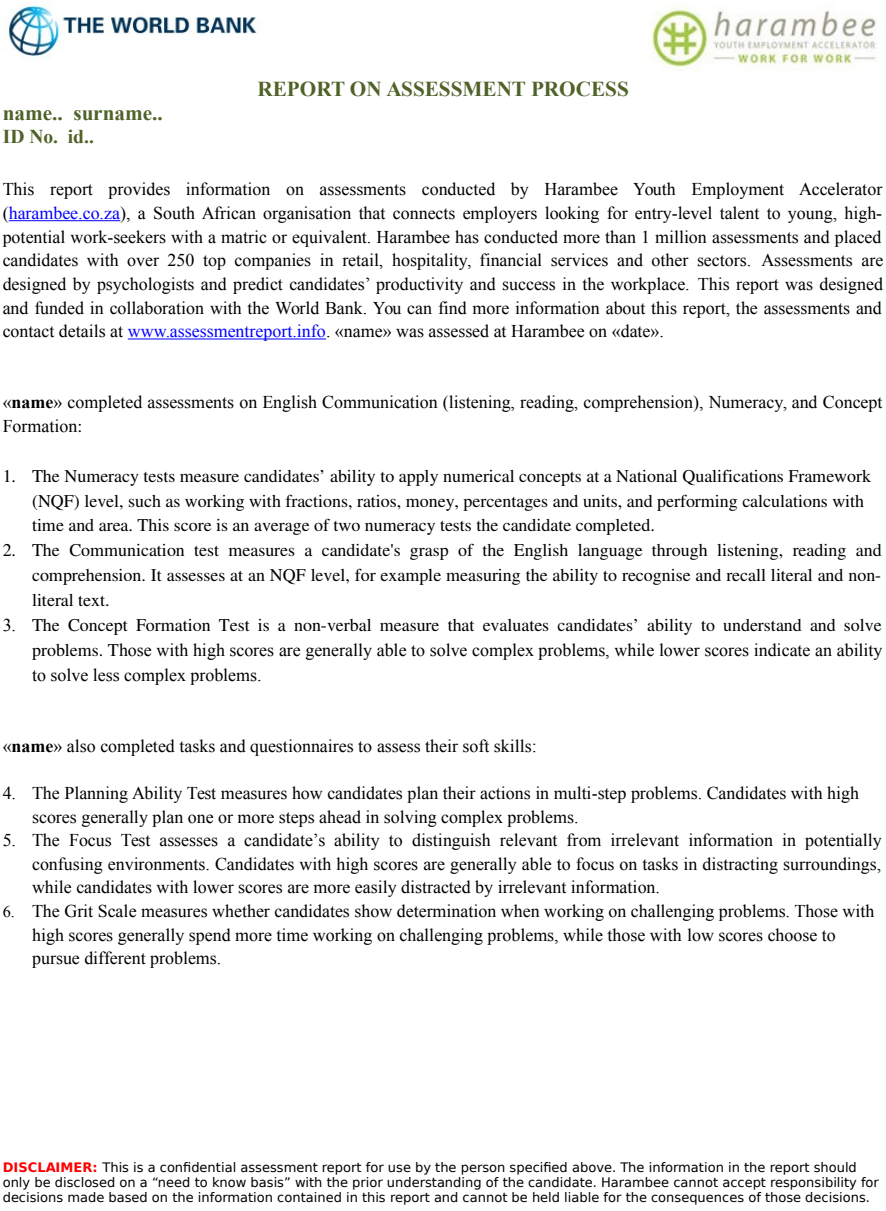
	Mean	Std Dev.	# Obs
Panel A: Characteristics of applications submitted			
Had one report in a vacancy with one report	0.12	0.33	3992
Had one report in a vacancy with three reports	0.38	0.48	3992
Had no report in a vacancy with one report	0.37	0.48	3992
Had no report in a vacancy with three reports	0.13	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 $\approx$ 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

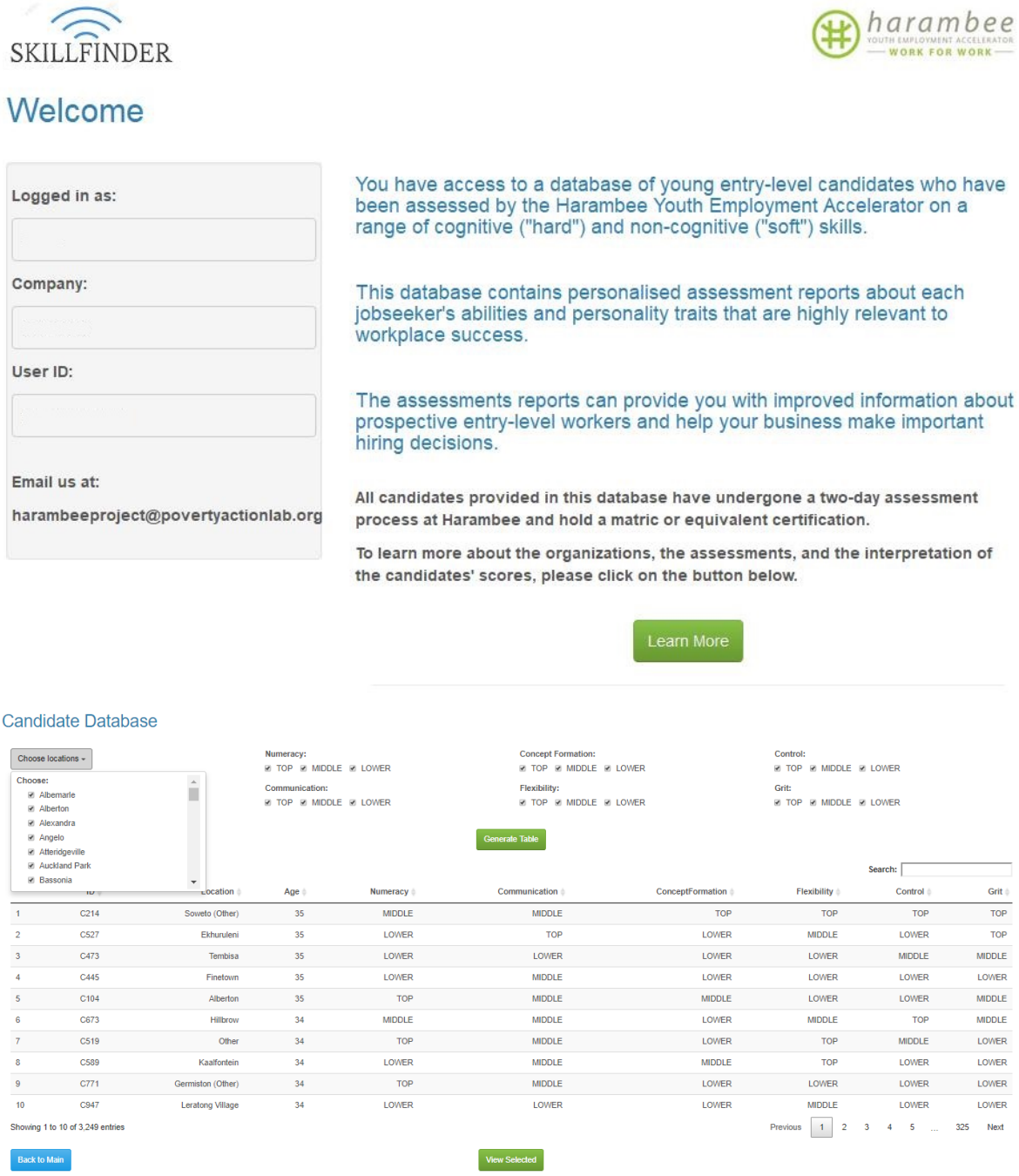


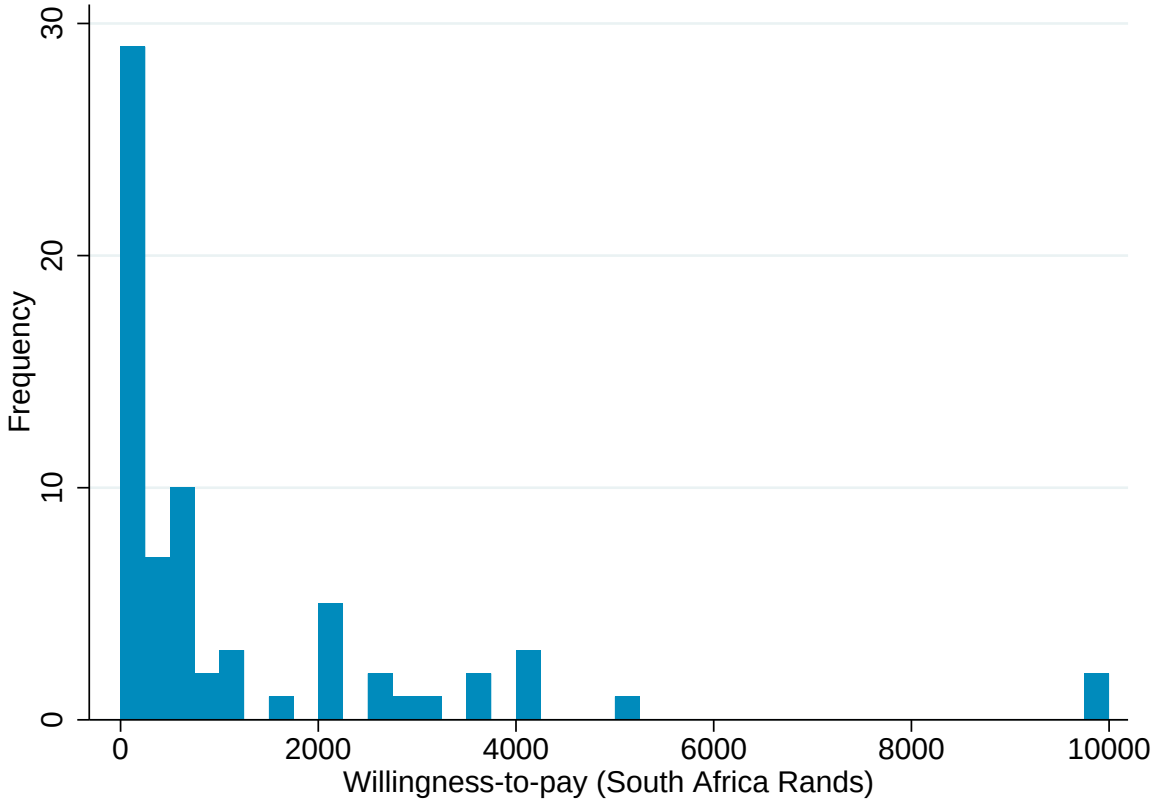
Figure G.2: Screenshot of Individual Candidate Profile on Platform

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

		(1)	(2)	(3)
Profile content		Share of firms ranking profile		Median
Top tercile	Highest education	First	Last	ranking
Communication	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.