Pre-Analysis Plan: Beliefs about Skills and Job Search Behavior

Robert Garlick   Lukas Hensel

Andrea Kiss   Kate Orkin

August 31, 2022

This is an analysis plan for studying how jobseekers' beliefs about their skills influence their job search decisions using a lab-in-the-field / framed field experiment. It builds on a completed field experiment registered at https://www.socialscienceregistry.org/trials/1631. This PAP is lodged after data collection has started but before any treatment effect analysis was conducted.

1   Experimental design

We recruit participants through our partner, SAyouth.mobi, a South African online platform run by the Harambee Youth Employment Accelerator where firms list vacancies for jobs and jobseekers can find resources to support job search. Our sampling frame for this platform consists of active platform users (i.e. active jobseekers) with addresses listed within commuting distance of our field site in Johannesburg. We invite sampled users to take part in a day of activities:

1. Jobseekers arrive, complete a consent process, and respond to a pre-treatment survey.

2. Jobseekers take assessments of their numeracy and communication skills and complete a brief survey about their perceived performance on the assessments.

3. **[Treatment group only]** Jobseekers receive a report on their performance in the numeracy and communication assessment. The report displays their performance in quintiles relative to the distribution of skills in a large sample of jobseekers from Johannesburg from similar backgrounds. They watch a video explaining the content of the report and have a chance to ask questions.

4. **[Control group only]** Jobseekers in the control group watch a video explaining the nature of assessment they took and have a chance to ask questions to the invigilators. They will receive the same report and explanation of the report but after the study is completed.

5. Jobseekers complete a survey covering their beliefs about their skills and job search plans.

6. Jobseekers make a series of 11 choices between two jobs. Each job pair has a relatively numeracy-focused and a relatively communication-focused job (based on expert ratings). The order in which job pairs are displayed is randomized at the individual level. For a subset of 5 job pairs we ask detailed beliefs about each job opportunity. For the last two randomly picked job pairs, each respondent receives information explaining which is the numeracy and which is the communication job.

7. Jobseekers make incentivised choices over different goods related to job search. One of these pairs contains two real jobs to which we submit applications on behalf of jobseekers. The other ten pairs contain adverts from real but not necessarily current jobs posted on SAyouth.mobi.

8. Jobseekers have the option of writing a cover email for the job they apply to for real. We use the time they spend writing as a proxy for job search effort.

9. We measure jobseekers' engagement on our partners' online platform over the next 30 days. This provides measures of search effort and potentially search targeting.

Randomization into treatment and control groups takes place at the fieldwork-day level. We create blocks of four sequential fieldwork-days and randomly assign two days in each block to treatment and two to control. The experiment started in August 2022 and is planned to finish in October 2022.

2   Hypotheses

We organize the analysis around two economic hypotheses. Under the first hypothesis, treatment shifts jobseekers' beliefs about their comparative advantage in communication versus numeracy skills, hence shifts their beliefs about their probability of getting jobs that require specific skills and wages conditional on getting those jobs, and hence affects the types of jobs to which they direct applications. Under the second hypothesis, treatment shifts jobseekers' beliefs about the level of their skills, hence shifts their beliefs about their probability of getting a job and wages conditional on getting a job, and hence affects the level of search effort they exert. These hypotheses are not mutually exclusive.

3   Jobseeker × job choice analysis

We pre-specify all equations with control variables and will report if results are sensitive to including controls.

Search targeting analysis   We will estimate the following specification:[1]

$$target_{icr} = \beta_1 treat_i + \delta_c + \delta_r + \delta_b + \beta X_i^{jc} + \varepsilon_{icr} \tag{1}$$

where $target_{icr}$ is a dummy variable equal to one if the job choice for job pair c in round r is aligned with individual i's comparative advantage in skills. We define comparative advantage in numeracy as being in a higher quintile in the numeracy test compared to the communication tests. Similarly, we define comparative advantage in communication as being in a higher quintile in the communication test compared to the numeracy tests.[2] For each job pair, we define a numeracy and communication job based on expert ratings (see Table 1 for an overview of the job pairs). $target_{icr}$ takes the value of one iff they choose the job that aligns with their comparative advantage in skills. $treat_i$ is a treatment indicator. δ_c are job pair fixed effects. δ_r are round fixed effects to address potential time trends within the survey. δ_b are randomization block fixed effects to address potential time trends across treatment days. The vector of individual level controls (X_i^{jc}) will contain age; female dummy; dummies for having a high school leaving certificate, having a post-secondary certificate, and for having a post-secondary degree; dummies for each of the skill quintiles for both numeracy and communication skills; and dummies for having a comparative advantage in each of the two skills.

Standard errors will be clustered at the fieldwork-day level (the level of treatment assignment).

Numeracy job title	Communication job title
Receiving and Dispatching Clerk	Sales Agent
Sales Teller	Customer service agent
Stock controller	General administrator
Laundry assistant	Waiter / waitress
Cashier	Host/hostess
Data capturer	Front desk assistant
Restaurant till manager	Receptionist
Store Cashier	Sales Assistant
Cash Teller	Recruitment administrator
Banking Call Center Agent	Retail Call Center Agent
Petrol Attendant	Maintenance Assistant

Table 1: Job titles used in the job choice task

Treatment effects on beliefs about jobs   To better understand why individuals adjust their targeting of job search, we measure beliefs about 10 of the 22 different jobs we ask about. We estimate treatment effects on beliefs using

---

[1]We will report if the effects on targeting differ by skill domain.

[2]We may also explore the joint relationship between treatment, measured skills, perceived skills, and targeting as an extension to test specific theoretical explanations.

$$bel_{icj} = \beta_1 treat_i + \beta_2 treat_i \times aligned_{icj} + \delta_r + \delta_j + \delta_b + \beta X_i + \varepsilon_{icj} \tag{2}$$

where bel_{icj} is one of three beliefs outcomes listed in Table 2 for individual i's belief about job j in job pair c. $aligned_{icj}$ is a dummy variable indicating that job j's skill profile (in terms of numeracy and communication skill requirements) aligns with an individual's comparative advantage in skills. The vector of individual level controls (X_i) will contain age; female dummy; dummies for having a high school leaving certificate, having a post-secondary certificate, and for having a post-secondary degree; dummies for each of the skill quintiles for both numeracy and communication skills; and dummies for having a comparative advantage in each of the two skills.

Concept	Notes
Desirability	Subjective rating from 0 to 10.
Expected wage	Using IHS transform to account for outliers.
Likelihood of getting job-offer	Ranging from 0 to 10.

Table 2: Belief outcomes about job opportunities

4   Jobseeker level analysis

We will estimate treatment effects on range of further outcomes at the individual level. For these outcomes we will estimate:

$$y_i = \beta_0 + \beta_1 treat_i + \delta_b + \beta X_i^{ind} + \varepsilon_i \tag{3}$$

where y_i is the outcome of interest for individual i. The vector of individual level controls (X_i^{ind}) will contain age; female dummy; dummies for having a high school leaving certificate, having a post-secondary certificate, and for having a post-secondary degree; dummies for each of the skill quintiles for both numeracy and communication skills; dummies for having a comparative advantage in each of the two skills; and a pre-treatment value of the outcome y_i where available.

We organize outcomes into families of conceptually similar measures, corresponding to steps in the two economic hypothesis described above. Within each family we designate some outcomes as primary and some as secondary. The primary outcomes are either the most direct measure of each concept or are indices that combine the secondary outcomes. All indices will be constructed following [Anderson, 2008]. In families where we have multiple primary outcomes, we will adjust for multiple testing by reporting sharpened q-values following Benjamini et al. [2006].

- **Beliefs about skills**[3]

---

[3]We will also analyse whether belief updating differs for individuals who are over- versus underconfident at baseline.

- Primary outcomes

  * Dummy indicating "accurate" beliefs about comparative advantage.[4]
  * Dummy indicating "accurate" beliefs about numeracy quintile.
  * Dummy indicating "accurate" beliefs about communication quintile.

- Secondary outcomes

  * Average level of confidence in skills (across numeracy and communication).[5]
  * Level of confidence in numeracy skills.
  * Level of confidence in communication skills.

- **Job search targeting**

  - Primary outcome: Number of planned job applications in the next 30 days to jobs that mainly require the skill in which the participant has a comparative advantage, minus the number for the other skill.[6]

- **Search effort**

  - Planned job search effort
    * Primary outcome
      · Total number of applications planned for the next 30 days.

  - Realized job-search effort
    * Primary outcomes
      · Index of secondary, platform-based outcomes
      · Time spent writing cover email as part of the field day (IHS transform).
    * Secondary outcomes based on activity on our partner's online job-search platform
      · Number of days active on platform in the 30 days following the treatment.
      · Number of job opportunities viewed in the 30 days following the treatment.
      · Number of job opportunities applied to in the 30 days following the treatment.[7]

- **Employment expectations**

  - Primary outcome
    * Index of all secondary outcomes

---

[4]We define accurate beliefs as beliefs that align with our skill measurements.

[5]We define confidence in beliefs as $skill\_belief_{ij} - skill\_actual_{ij}$, where $skill\_belief_{ij}$ is individual i's belief about their skill j and $skill\_belief_{ij}$ is individual i's actual skill level (both measured in quintiles).

[6]We are also exploring whether we can obtain data on search targeting of jobseekers on our partner's platform.

[7]We count all clicks on the "apply" button as applications whether or not they complete the application. We will observe actual applications for a subset of job opportunities and will study them as a robustness check.

- Secondary outcomes
    * # expected call backs / total planned application
    * # expected job offers / total planned per application
    * Expected search duration for formal job (IHS transformed, reverse coded in the index)
    * Probability of being employed one month after treatment.
    * Probability of being employed three months after treatment.

- **Wage expectations**

    - Primary outcome
        * Index of all secondary outcomes
    - Secondary outcomes
        * Expected earnings in formal job (IHS transform)
        * Minimum expected earnings in formal job (IHS transform)
        * Maximum expected earnings in formal job (IHS transform)
        * Reservation wage (IHS transform)[8]

- **Wage and employment expectations by job skill requirements**

    - Primary outcomes
        * Difference between expected number of offers from / expected number of applications to jobs that mainly require skill X and jobs that mainly requires skill Y
        * Difference between expected months to get a job requiring mainly skill Y and a job requiring mainly skill X
        * Difference between expected wage in a job that mainly requires skill X and a job that mainly requires skill Y

    where X and Y are skills in which participant respectively has and does not have a comparative advantage.

- **Willingness-to-pay**

    - WTP for numeracy learning materials.
    - WTP for communication learning materials.
    - WTP for information about the relative skill requirements for the job choices.

---

[8]The reservation wage is a decision rule in standard job search models, derived partly from the wage distribution. This is not inconsistent with our interpretation of reservation wages as another proxy for the expected wage distribution.

5   Additional analysis

Heterogeneity analysis   We may also conduct exploratory heterogeneity analysis informed by the economic hypotheses. In particular, the hypotheses suggest that treatment effects depend on participants' baseline beliefs about the level of their skills and their comparative advantage.

Decoding skill requirements   To test whether jobseekers can decode the skill requirements from the vacancies provided as part of the experiment, we ask jobseekers to assess the communication and numeracy skill requirements for 5 of the 11 job pairs. We then calculate the average fraction of correctly decoded job adverts. We further analyze decoding by beliefs about comparative advantage and search targeting. If we find that jobseekers have very different beliefs about the skill content of jobs than our assessment, we will analyze how this affects their job targeting.

In addition to this, we will test whether revealing the skill requirements affects jobseekers' targeting behavior. For this purpose, we will reveal the relative skill requirement for the last two job pairs jobseekers are asked about.[9] We then estimate the following equation to see whether revealing the skill requirement affects jobseekers' search targeting decision:

$$target_{icr} = \beta_1 treat_i + \beta_2 treat_i * revealed_{ic} + \delta_c + \delta_r + \beta X_i + \varepsilon_{icr} \qquad (4)$$

β_2 will indicate whether revealing the skill demand affects jobseekers targeting decision.

Impact on variance of beliefs   For some of the skill and wage beliefs, we will elicit a full distribution of beliefs (rather than just the expected value). For those variables, we will also estimate treatment effects on the estimated variance of beliefs to assess whether the treatment reduces uncertainty in beliefs.

References

Michael L. Anderson. 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(484):1481–1495, 2008. ISSN 0162-1459. doi: 10.1198/016214508000000841.

Yoav Benjamini, Abba Krieger, and Daniel Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006.

---

[9]We randomize the order in which job pairs are presented, so this is equivalent to revealing skills for two random job pairs.