# Rebuild: Refugee Economic and Social Integration in East Africa Pre-Analysis Plan

Travis Baseler* Ibrahim Kasirye† Thomas Ginn‡ Belinda Muya§
Andrew Zeitlin¶

Version: May 2, 2023

## 1 Overview of project

Micro-entrepreneurship is a common form of employment for urban residents, but many small firms yield low profit. Borrowing constraints (de Mel et al., 2008) and a lack of managerial capital (Brooks et al., 2018) may act to constrain business growth. These constraints are potentially amplified for those with limited access to existing business networks like refugees and women.

Mentorship by an experienced business owner and peer groups of other inexperienced business owners has the potential to grow existing business networks. Mentors may have valuable information to share and can serve as role models to help build confidence. Further, a program intervention may facilitate exchange across groups that would otherwise have few connections, such as across gender or nationality lines. This contact may also increase intergroup social cohesion, a hypothesis studied by social scientists for decades. However, the mentor's effort is likely difficult for mentees or program facilitators to observe, and the mentor may view the mentees as potential competitors, reducing their incentive to invest in mentorship and share valuable information.

This project tests whether a group business mentorship program can improve microenterprise success and social cohesion in Kampala, Uganda, a city that hosts nearly 100,000 refugees. We randomly vary business grants, inclusion in a mentorship group, the composition of groups along gender and nationality lines, and a "shared fate" component that compensates all group members for the success of their partners' businesses."Shared fate" models may potentially better align incentives by giving mentors a stake in the mentees' business success. This model may also increase social cohesion in heterogeneous groups by inducing a stronger sense of shared purpose.

The study is designed to address the following overall research questions:

1. What is the effect of the cash grants, cash grants with mentorship, and cash grants with mentorship in the "shared fate" model on economic and social cohesion outcomes?

   (a) How do these effects differ between hosts and refugees?

*University of Rochester, travis.baseler@rochester.edu
†Economic Policy Research Centre, kasiryeibra@hotmail.com
‡Center for Global Development, tginn@cgdev.org
§International Rescue Committee, belinda.muya@rescue.org
¶Georgetown, andrew.zeitlin@georgetown.edu

(b) How do these effects differ between men and women?

2. What is the effect of being assigned to a mixed mentorship group (either mixed refugee and host or mixed genders) compared to a homogeneous mentorship group on economic and social cohesion outcomes?

   (a) How do these effects differ between hosts and refugees?

   (b) How do these effects differ between men and women?

# 2 Project design and sample

## 2.1 Sample

This experiment takes place with a projected sample of 2,000 inexperienced or prospective micro-entrepreneurs living in Kampala, Uganda, and 650 mentors. The sample includes both men and women, and both Ugandan nationals and refugees. Study participants were actively recruited from relevant populations in Kampala. Study inclusion criteria were as follows. Among mentees, study inclusion criteria are:

- Age less than or equal to 35 years;
- Speak at least one of Luganda or English at a conversational level;
- Fewer than 6 years of business experience

Among mentors, study inclusion criteria are:

- Age greater than or equal to 25 years;
- Speak at least one of Luganda or English at a conversational level;
- Four or greater years of business experience;
- Judged by IRC staff to have sufficient business skills to be an effective mentor.

## 2.2 Interventions

We are implementing seven treatments in partnership with the International Rescue Committee (IRC).

**Treatment 1: Grant** In this treatment arm, participants receive a cash grant of 2,000,000 UGX (about US$524). The grant is labeled as being intended for business purposes, but there is no spending oversight.

**Treatment 2: Grant + Homogenous Mentorship Group** This treatment arm combines a grant with an invitation to participate in a business mentorship group. Each group contains four members—one mentor and three mentees—of the same gender and nationality.

**Treatment 3G: Grant + Gender-Heterogeneous Mentorship Group** This treatment arm combines a grant with an invitation to participate in a business mentorship group of four members including two men and to women of the same nationality.

**Treatment 3N: Grant + Nationality-Heterogeneous Mentorship Group** This treatment arm combines a grant with an invitation to participate in a business mentorship group of four members including two Ugandans and two refugees of the same gender.

**Treatments 2', 3G', and 3N': Grant + Mentorship Group with Shared Fate** These treatment arms modify arms 2, 3G and 3N by introducing a "shared fate" model which rewards group members based on the success of other members' businesses.

**Control Group** This group was told that they would be eligible for the program 18 months in the future, but receives no treatment during the course of this experiment.

All treatment groups, not including control, received information about refugee hosting following the design of Baseler et al. (2022).

### 2.3 Randomization

Study participants were assigned to treatment arms in a stratified, randomized assignment. This proceeds in the following steps:

1. Study participants are first assigned to strata of 52 individuals. Comprising a single nationality, these contain 40 potential clients and 12 potential mentors, with half of mentors and mentees female. Strata are clustered geographically to the extent possible.

2. Strata are paired, to comprise two-strata pairs in which one strata consists of refugees and one strata consists of Ugandan nationals. Within such strata pairs, all permutations of mentorship groups within the study design can be formed.

3. Individuals within strata pairs are then assigned to treatments. For mentees, assignment probabilities and project sample sizes (in parentheses) are:

    - Control: 0.175 (350)
    - T1 Grant: 0.225 (450)
    - Mentorship ("basic"):
        - T2 Homogenous identity: 0.15 (300)
        - T3G Heterogeneous identities, by gender: 0.1125 (228)
        - T3N Heterogeneous identities, by refugee status: 0.1125 (222)
    - Mentorship ("shared fate"):
        - T2' Homogenous identity: 0.075 (150)

  - T3G' Heterogeneous identities, by gender: 0.075 (150)
  - T3N' Heterogeneous identities, by refugee status: 0.075 (150)

4. For mentors, projected sample sizes are:

   - Control: 200
   - Mentorship ("basic"):
     - T2 Homogenous identity: 100
     - T3G Heterogeneous identities, by gender: 76
     - T3N Heterogeneous identities, by refugee status: 74
   - Mentorship ("shared fate"):
     - T2' Homogenous identity: 50
     - T3G' Heterogeneous identities, by gender: 50
     - T3N' Heterogeneous identities, by refugee status: 50

To facilitate subsequent randomization inference, we collect 2,000 permutations of assignments meeting a criterion for balance on a vector of attributes observed in IRC registration data. The final assignment is chosen at random from among this collection of feasible randomizations. The feasible randomizations and the final assignment selections were all generated in R.

Balance checks will be run on baseline data to confirm the random nature of the assignment process. Feasible randomizations will meet a criteria defined in absolute terms for the balance of a set of attributes observed in IRC registration data.

### 2.4 Timeline

Enrollment and baseline for this study began over the summer of 2022. The following data collection rounds are planned for post-intervention measurements :

- 3-month midline: Nov 2022–Jan 2023
- 6-month midline: Mar–Apr 2023
- 9-month midline: July–Aug 2023
- 12-month midline: Nov–Dec 2023

Due to a delay in program roll-out, approximately 400 participants were treated in a second wave in January 2023, with follow-ups scheduled for April, July, and October of 2023, and January of 2024. We will include survey-round, calendar-month, and randomization stratum fixed effects in all estimation models to account for the staggered roll-out.

## 2.5 Hypotheses and Outcomes

Our outcomes of interest are organized into conceptual domains, such as "social cohesion" and "business success." Within each domain we compute a summary index, using the approach of Anderson (2008). Our analysis will be organized into three sets of tests, which we approach differently:

1. Primary hypotheses about summary indices, which we test individually as well as jointly.
2. Secondary hypotheses about summary indices, which we test individually.
3. Outcomes on the full set of index components, for which we report sharpened q-values within each domain.

For tests that pool across treatment groups, we will note this explicitly by referring to the treatment arms being pooled in parentheses. For example, all pooling all mentorship groups will be denoted as (T2,T2',T3G,T3N,T3G',T3N'). For shorthand, we use T3 to refer to the pooled set of basic heterogeneous mentorship groups (T3G,T3N), and T3' to refer to the pooled set of shared-fate heterogeneous mentorship groups (T3G',T3N').

### 2.5.1 Outcome Domains

Unless otherwise specified, each question is measured with a 5-point Likert scale and an option for "don't know" (which is coded as missing).

***Domain 1: Business success***

- A binary indicator for whether the respondent operated a business in the past month.
- Self-reported profits from all businesses over the past month. We will code zeros for respondents without operational businesses.

***Domain 2: Social cohesion***

- Social proximity, computed as an Anderson (2008) index over:
  - "I would only be comfortable marrying a refugee/Ugandan, not someone of another nationality."
  - "I would only be comfortable having a refugee/Ugandan marry a member of my family, not someone of another nationality"
  - "I would only be comfortable having a refugee/Ugandan as a close, personal friend, not someone of another nationality"
  - "I would only be comfortable having a refugee/Ugandan as a neighbor, not someone of another nationality"
- "When [refugees'/Ugandans'] businesses are successful, [Ugandans/refugees] benefit"

- Inter-group altruism, measured by the amount donated to an anonymous receiver of a different identity in a dictator game.
- Inter-group trust, measured by the amount of money sent to an anonymous partner of a different identity in an incentivized trust game.
- Inter-group trust, measured by willingness to make a job referral to a confederate staff member of another nationality.
- "Would you be open to collaborating with business owners from a country other than [origin country]?"

***Domain 3: Business practices***

- A binary variable indicating whether the business is officially registered.
- Business capital. We sum two variables.[1]
  - "If you were to sell all the [business equipment] you own right now, how much do you think you could make?" This is asked item-by-item and summed.
  - "If you were to sell all the inventory you own right now, how much do you think you could make?"
- Number of hours worked at all businesses over the past 7 days.
- "How much total business-related debt do you currently have?"
- Number of contacts listed in a business networks module.
- The number of "Yes" responses to the following questions about whether they did the following in the past 30 days:
  - Visited at least one of your competitor's businesses to see what prices your competitors are charging.
  - Visited at least one of your competitor's businesses to see what products your competitors have available for sale.
  - Asked existing customers whether there are any other products the customers would like the business to sell or produce.
  - Talked with at least one former customer to find out why former customers have stopped buying from your business.
  - Asked a supplier about which products are selling well in your business' industry.
  - Attracted customers with a special offer on price.
  - Attempted to negotiate with a supplier for a lower price on supplies.

[1] We analyze the inverse hyperbolic sine transformation of the total.

  - Compared the prices or quality offered by alternate suppliers or sources of raw materials to your business' current suppliers or sources of raw material.
  - Ran out of stock or raw materials once per month or more. ("No" is counted as 1.)
  - Record every purchase and sale made by your business.
  - Kept a complete written budget, which states how much is owed each month for rent, electricity, supplies, and all other costs to business.

- "Over the last 30 days, how often did you spend money advertising your business? Every day, every week, every month, a couple times, or never?"
- "How often did you keep written books/accounting records? Always, frequently, sometimes, occasionally, or never?"
- "How often did you sell goods or provide services to customers on credit? For all sales, most sales, some sales, a few sales, or never?" (will be excluded from domain index)
- "How often did you buy materials, tools, or machines for your business on credit? For all sales, most sales, some sales, a few sales, or never?" (will be excluded from domain index)
- "In the last 3 months, have you offered a new product or service at your business that you did not offer previously?"
- "In the past 30 days, have you thought carefully about what your specific goals for your business are in the next 6 months or so?"
- Do you keep your business and personal finances separate?
- Do you feel like an outsider in the small business community?
- Do you feel like you have the basic skills and abilities to be successful in business?
- Do you feel like you will be able to overcome the difficulties you experience in business?
- How confident do you feel about successfully identifying new business opportunities?

***Domain 4N: Inter-Nationality Contact***

- "Think about all of your business' customers. How many of your customers are from a different country than you? All, most, some, few, or none?"
- "How many of your business collaborators are from a country other than [origin country]?"
- "How many of your suppliers are from a country other than [origin country]"
- "Are any of your employees from a different country than you?"
- "In the past 30 days, how many [refugees/Ugandans] have you contacted for any social reason, such as having a long conversation?"

- Number of people from another country listed in the networks module.
- "How often do you have any contact with [other nationality] when you are out? This could be on public transport, in the street, in shops or in the neighbourhood."
- Indicator for whether they participate in any social activities with members of the other nationality.
- "Please think about the businesses you have sold goods or services to within the past 3 months. Do not count consumers, just other businesses. How many of those businesses are managed by people from [other nationality]"

***Domain 4G: Inter-Gender Contact***

- Number of people from another gender listed in the networks module.

***Domain 5: Household Well-Being***

- Total household income. This will be computed as the sum of 4 measures:
  - "What were the profits of your business during the last 30 days?"
  - "What were the profits of [any other household-owned] businesses (excluding this one) during the last 30 days?"
  - "How much wage and salary income did you earn in the last 30 days?"
  - "How much wage and salary income did [other members of your household] earn in the last 30 days?"[2]
- Total household savings, added across sources (microfinance, SACCO, bank, cash, informal, crops, mobile money, other).
- Change in total value of household durables over last 3 months, asked item-by-item, and computed as purchases – sales.
- Business capital (see D3).
- Total value of household debt (enters the index negatively), computed as the sum of debt owed to friends/neighbors, relatives, private money lenders, coops/associations, agricultural input suppliers, SACCOs, banks, microfinance organizations, NGOs, other debt, and outstanding school fees and medical bills.
- "Compared to the average Ugandan in Kampala, how would you describe the economic situation of your household? Much better, somewhat better, about the same, somewhat worse, or much worse?"
- "Over the past 7 days, how many days did someone in your household skip a meal because you didn't have enough money for food?"

[2]This includes wages paid by businesses owned by the household, which are otherwise not included in profits.

- "Over the past 30 days, how often have you or anyone in your household struggled to afford basic household expenses (such as medicine, rent, school fees)?"
- "In the past 30 days, have you or anyone in your household had to sell assets (jewelry, furniture, clothing, tools, machines, land) in order to afford basic household expenses?"
- "How many months in the last 3 months have you not been able to pay rent for your home?"
- "If your household had an emergency that required 50,000 UGX urgently (within 3 days), would you be able to find the money?"
- Over the last 3 months, how many children in your household between the ages of 6 and 17 have missed school for at least 1 month?
- In the last 30 days, how many days did someone in your household miss work because of a physical or psychological health condition?
- "In the past 3 months, have any children in the household under the age of 15 worked because of a lack of cash for your family to meet their basic needs?"

***Domain 6: Women's Bargaining Power***

- Decision-making power in household spending decisions (ranked from most to least, with "not applicables" in the middle, then treated as a Likert)
- Has input in decision to work
- Has input in type of work
- Has input in childbearing
- Has input in children's education decisions
- An indicator for whether male adults are ranked strictly above female adults in the order that household members eat when food is in short supply (enters negatively).
- "In your opinion, is a husband justified in hitting or beating his wife if she burns the food?" (enters negatively)
- "In your opinion, is a husband justified in hitting or beating his wife if she neglects the children?" (enters negatively)

***Domain 7: Psychological Well-Being***

- "In the past month, how much of the time were you a happy person?"
- "In the past month, how much of the time did you feel calm and peaceful?"
- "In the past month, how much of the time did you feel down-hearted and sad?"

***Domain 8: Social Proximity for Refugees***

- "How connected do you feel with Uganda?"
- States that they plan on living in Uganda in the future.
- "How often do you feel isolated from Ugandan society?"
- "In the last 12 months, how often did you eat dinner with Ugandans who are not part of your family?"
- Participates with Ugandans in job-related groups.
- Participates with Ugandans in other hobbies.
- Participates with Ugandans in religious groups.
- "In the last 12 months, how often have you provided such everyday favors to Ugandans, such as lending items, borrowing a little money, or watching children?"
- "In Uganda, how difficult or easy would it be for you to see a doctor?"
- "In Uganda, how difficult or easy would it be for you to search for a job with a salary?"
- "In Uganda, how difficult or easy would it be for you to get help with legal problem?"

***Domain 9: Beliefs about economic effects of refugees***

- "How do the [sector] businesses managed by people from other countries affect your business overall? Do they help you a lot, help you a little, hurt you a little, hurt you a lot, or have no effect on you?"
- "Taking everything into consideration, would you say the overall economic effect of refugees on Uganda has been positive, negative, or neutral?"
- "How about the overall economic effect of refugees on you personally?"
- "How many refugees have skills and contribute to the economy?"

***Domain 10: Support for inclusive refugee hosting.***

- "Refugees in Uganda should be required to live in the settlements. Refugees should not be allowed to live in Kampala."
- "Refugees in Uganda should not be allowed to work outside the settlements."
- "Uganda should not provide land for farming to refugees in the settlements."
- "Refugees in Uganda should not be allowed to vote in Uganda or become full Ugandan citizens."
- "Uganda should not accept more refugees."

### 2.5.2 Primary Hypotheses

Below, we express hypotheses as signed hypotheses, however, we conduct two-sided tests against the relevant (sharp) null of no effect.
We have four main hypotheses related to two primary outcomes: business success and social cohesion. Our four hypotheses are:

- Mentorship with a grant (T2,T2',T3,T3') will improve business success (D1) relative to a grant alone (T1).
- The shared fate model (T2',T3') will improve business success (D1) relative to a fixed mentor payment (T2,T3).
- Heterogeneous mentorship (T3,T3') will improve business success (D1) relative to homogenous mentorship (T2,T2').
- Nationality-heterogeneous mentorship (T3N,T3N') will increase social cohesion (D2) relative to other mentorship groups (T2,T2',T3G,T3G').

### 2.5.3 Secondary Hypotheses

We will test the following additional hypotheses:

- Among nationality-heterogeneous groups, the shared fate model (T3N') will increase social cohesion (D2) relative to fixed-pay mentorship (T3N).
- Mentorship (T2,T3,T2',T3') will improve business practices (D3) relative to a grant alone (T1).
- The impact of the shared fate treatment for business success (D1) and social cohesion (D2) will be greater in heterogeneous groups than homogenous groups (T3'–T3 > T2'–T2).
- Nationality-heterogeneous mentorship groups (T3N,T3N') will increase inter-nationality contact (D4N) relative to other mentorship groups (T2,T2',T3G,T3G').
- Gender-heterogeneous mentorship groups (T3G,T3G') will increase inter-gender contact (D4G) relative to other mentorship groups (T2,T2',T3N,T3N').
- Any mentorship (T2,T3,T2',T3') will increase household well-being (D5) relative to cash alone (T1).
- Cash (T1) will increase business success (D1) and household well-being (D5) relative to control (C).
- Any treatment (T1,T2,T3,T2',T3') will increase bargaining power (D6) for female participants relative to control (C).

- Any treatment (T1,T2,T3,T2',T3') will increase psychological well-being (D7) relative to control (C).
- Nationality-heterogeneous mentorship (T3N,T3N') will increase social proximity for refugees (D8), relative to other mentorship groups (T2,T2',T3G,T3G').
- Nationality-heterogeneous mentorship (T3N,T3N') will increase the valence of Ugandan participants' beliefs about the economic effects of refugees (D9), relative to other mentorship groups (T2,T2',T3G,T3G').
- Nationality-heterogeneous mentorship (T3N,T3N') will increase support for inclusive refugee hosting (D10) among Ugandans, relative to other mentorship groups (T2,T2',T3G,T3G').
- The grant alone (T1, which includes information about refugee hosting) will increase support for inclusive refugee hosting (D10) and social cohesion (D2) relative to control (C).
- Nationality-heterogeneous mentorship (T3N,T3N') will increase perceptions about the other nationality's business success relative to other mentorship groups (T2,T2',T3G,T3G'). This outcome is measured by the question, "Think about small businesses in your area that are new—that is, they were opened recently by a young, inexperienced owner—or that could be started soon. Out of 10 small businesses like this operated by a [demographic], how many do you think will survive for at least one year?"
- All treatments (T1,T2,T3,T2',T3') will increase knowledge about refugee hosting among Ugandans, relative to control (C). We will measure knowledge using an Anderson (2008) index over two questions:
  - "Are any of the international donations to refugees in Uganda shared with Ugandans?"
  - "Are refugees allowed to live outside of the camps or settlements?"
- Nationality-heterogeneous mentorship (T3N,T3N') will increase Ugandans' perceptions about whether refugees have skills and contribute to the economy, relative to other mentorship groups (T2,T2',T3G,T3G').
- Nationality-heterogeneous mentorship (T3N,T3N') will increase Ugandans' perceptions about whether refugees can support themselves without assistance, relative to other mentorship groups (T2,T2',T3G,T3G').
- We will test for differences in mentorship group meeting frequency, the number of 1–1 interactions outside of group meetings, and the aspects of group dynamics (share of members reporting that the group's discussions are relevant for their business, that the group has helped connect them with new customers or suppliers, and that everyone in the group is working hard at their business and contributing to the group) in (T2,T2') vs (T3G,T3G'), (T2,T2') vs (T3N,T3N'), (T2'T3') vs (T2,T3), and (T3'-T3) - (T2'-T2).

- Mentorship will improve business success (D1) for mentors, relative to control (C).
- The number of times that a mentorship group wins the lottery, conditional on the number of times it was eligible for the lottery, will increase social cohesion (D2).
- As a potential mechanism for differences between heterogeneous and homogenous mentorship groups, or between shared fate and basic mentorship, we will test whether those groups differ in the following measures of intra-group dynamics (asked at the pair level for each pair combination in the group, and averaged across pairs in each group):
  - Perceived ability of other group members.
  - Perceptions of whether two members have a lot in common.
  - Whether two members shared customers (Yes/No).
  - Whether two members shared suppliers (Yes/No).
  - Whether two members shared business skills (Yes/No).

We will also test the following secondary hypotheses related to heterogeneous treatment effects:

- The impact of the grant (T1) on business success (D1) and social cohesion (D2) will be larger for refugees.
- The impact of nationality-heterogeneous mentorship (T3N,T3N') on social cohesion (D2) relative to other mentorship groups (T2,T2',T3G,TG3') will be different for Ugandans and refugees.
- Impacts of mentorship (T2,T3,T2',T3') on business success (D1) relative to grant alone (T1) will be larger for female and refugee business owners.
- The impact of mentorship (T2,T3,T2',T3') on business success (D1) relative to a grant alone (T1) will be greater for participants with smaller business networks at baseline (measured as the number of contacts in the network module).
- The impact of nationality-heterogeneous mentorship (T3N,T3N') on social cohesion (D2) relative to other mentorship groups (T2,T2',T3G,T3G') will be greater for participants with less inter-group contact at baseline.
- The impact of mentorship (T2,T3,T2',T3') on business success (D1) and social cohesion (D2) relative to grant alone (T1) will be greater when the mentor runs a more profitable business at baseline.
- The impact of mentorship (T2,T3,T2',T3') on business success (D1) and business practices (D3) relative to the grant alone (T1) will be larger among those with worse practices at baseline.

- We will test whether impacts of nationality-heterogeneous mentorship (T3N,T3N') on social cohesion among Ugandans, relative to other mentorship groups (T2,T2',T3G,T3G'), are greater for those with lower baseline beliefs about whether refugees have skills and contribute to the economy.

- The impact of gender-heterogeneous mentorship (T3G,T3G') relative to other mentorship groups (T2,T2',T3N,T3N') on business success (D1) will be greater for women than for men.

- The impact of nationality-heterogeneous mentorship (T3N,T3N') relative to other mentorship groups (T2,T2',T3G,T3G') on business success (D1) will be greater for refugees than for Ugandans.

- We will test whether business success (D1) impacts in heterogeneous groups (T3,T3') depend on whether the mentee and the mentor share the same gender or nationality.

- Within nationality-heterogeneous groups (T3N,T3N'), we will test whether business success (D1) among mentees of a different nationality than their mentor (relative to success among mentees of the same nationality as their mentor) are greater when the mentor has a higher prior perception of the opposite nationality's average business ability, conditional on their perception of the same nationality's average business ability.

Heterogeneous treatment effects will be estimated by including an interaction of the heterogeneity dimension and the treatment status in the main ANCOVA specification. When the dimension of heterogeneity is not a binary variable, we will convert it to a binary variable by splitting at the sample median.

# 3 Statistical analysis

## 3.1 Measurement

Indices will be constructed as in Anderson (2008). Index components with 90% or more of respondents providing a directionally positive response (after transformation from Likert to binary) in the control group will be dropped.

We will transform variables in the following ways:

- Likert scales and other categorical variables will be transformed into binary measures split around the median response, with the median resolved toward the smaller group. "Don't know" and other missing values will not be included in the index; we will not impute missing values for outcome variables. We will impute missing values for control variables using the baseline mean.

- Monetary measures including profit, revenue, capital, inventory, debt, & savings:

- Values will be winsorized at the 1st and 99th percentiles within each survey round by treatment group by nationality.
- Nominal values will be converted to real values using the CPI from the Uganda Bureau of Statistics.
- If the respondent declines to provide an exact amount but provides a range (pre-specified on the survey), we will use the midpoint of the range. For firms in the top range, the median of firms in the top range with reported point estimates will be used. For firms in the bottom range of measures that can take on negative values (such as profit), we will use the median of firms in the bottom range with reported point estimates.
- For firms that are not operating, values will be recorded as 0.

#### 3.1.1 Estimating samples

Mentors—who were randomized into either T2, T2', T3, T3', or a control group—will be included in regressions of Domains 2, 4, 8, 9, and 10. We will examine effects on mentors in other domains separately.

### 3.2 Testing and estimation

The starting point for all analyses of intent-to-treat effects will be one of two empirical models of outcomes. Our choice of model depends on the nature of the outcome under study.

For outcomes that may contain negative values (including profits, and all normalized indices), or for outcomes that are bounded from both above and below (such as binary variables), we will estimate intent-to-treat effects using the following ANCOVA specification:

$$y_{ist} = \sum_{j=1}^{5} \beta_j T_{ji} + \gamma y_{i0} + \delta M_{i0} + \eta X_i + \theta_t + \gamma_d + \alpha_s + \epsilon_{ist}, \tag{1}$$

where $y_{it}$ is an outcome for individual $i$ measured at time $t$ with $t = 0$ corresponding to baseline (pre-treatment) values, $M_{i0}$ is an indicator for a missing value of $y_{i0}$, $T_{ji}$ are treatment assignment dummies for treatment groups $j = \{1, 2, 3, 2', 3'\}$, $X_i$ is a vector of baseline controls chosen through double lasso, $\theta_t$ is a survey round fixed effect, $\gamma_d$ is a calendar date (month) fixed effect, $\alpha_i$ is a randomization stratum fixed effect, and $\epsilon_{it}$ is an error term. Standard errors will be clustered at the individual level. We will run separate lassos for each dependent variable using the Stata package *pdslasso* and include the following candidate control variables from the baseline in each: baseline values of all primary and secondary outcomes, in individual and summary-index form, where available, with quadratics of all continuous and unbounded variables; values of gender, age, and indicators for nationality as well as for languages spoken, all at registration. Baseline values of the outcome in question (as well as indicators for missingness of this variable), as well as strata, survey round, and calendar-month fixed effects will be partialled out for this exercise. All missing values at baseline will be imputed prior to inclusion in the lasso, and indicators for imputed values will be included as well as candidate covariates.

Where outcome variables are by construction weakly greater than zero and unbounded from above,[3] we will replace the model of equation (1) with an analogous Poisson Quasi-Maximum Likelihood Estimate (QMLE) specification:

$$\mathrm{E}[y_{ist}] = \exp\left\{\sum_{j=1}^{5} \beta_j T_{ji} + \gamma y_{i0} + \delta M_{i0} + \eta X_i + \theta_t + \gamma_d + \alpha_s\right\}. \tag{2}$$

### 3.2.1 Testing

Our primary approach to testing hypotheses of equality between pairs of treatment arms will be to use randomization inference, permuting only those treatments for which equality is implied by the specific hypothesis in question. We will use 2,000 such feasible randomiziations, each allowable under the assignment mechanism.

From each of these feasible randomizations, we will use one of two candidate test statistics as the basis for the distribution under the sharp null against which the actual assignment will be compared: a Kolmogorov-Smirnov (henceforth KS) statistic, or the relevant $t$ statistic from the appropriate outcomes regression in Equations (1) or (2) above.

Data from the control group will be used to decide between the two test statics. Specifically, we will use control-group data, pooling all time periods, to simulate 2,000 draws from a generative model (of sample size equal to our empirical sample). These draws will impose a value of the constant-effects treatment effect (e.g., the coefficient $\beta_j$) for the hypothesis in question that is equal to 0.2 standard deviations of the control-group distribution. We will use these simulated trials to compare the power between the two proposed testing approaches described below, and we will select the test statistic that has greater power against the sharp null hypothesis in this exercise.

Our first candidate test statistic will be the $t$ statistic, $t_j$, on the on the appropriate coefficient in the relevant outcomes regression (Equation (1) or (2)).

Our second candidate test statistic is a residualized $KS$ statistic, which we compute in the following steps:

1. To take advantage of the ability of regression controls to increase power, we first residualize outcome $y_{it}$ based on the ANCOVA outcomes regression in Equation (1) (or modify the below analogously for Equation (2) as appropriate). Using estimated coefficients from this regression we define the residualized outcome

$$\ddot{y}_{ist} = y_{it} - (\gamma y_{i0} + \delta M_{i0} + \eta X_i + \theta_t + \alpha_s) \tag{3}$$

2. For a hypothesized difference in outcomes between treatments $j, j' \in \mathcal{T} = \{T_1, T_2, T_3, T_2', T_3'\}$, our test statistic for randomization inference is the KS statistic, defined as

$$KS = \sup_{\ddot{y}} \left| \hat{F}_j(\ddot{y}) - \hat{F}_{j'}(\ddot{y}) \right| \tag{4}$$

[3] Examples include (but are not limited to) business revenues, productive asset values, and savings.

We calculate the value of this test statistic under the realized assignment.

3. This test statistic is compared to a reference distribution based on 2,000 permutation of treatments $j, j'$ that were feasible under the original assignment mechanism (i.e., those that pass the balance threshold used to select feasible treatments in the actual randomization).[4] The $p$-value for this test is then given by the share of draws in this reference distribution for which the associated KS test statistic is larger in absolute magnitude than the estimated KS statistic under the realized assignment.

For hypotheses regarding heterogeneous effects we use the appropriate outcomes regression in either Equation (1) or (2), augmented with an interaction term. We use asymptotic standard errors, clustered at the individual level, $i$.

For hypotheses that compare pooled groups of treatments, we use randomization inference in the relevant framework (KS or $t$ statistic), using the set of feasible randomizations that exchange pooled treatments, and estimating a specification that groups pooled treatments as appropriate.

#### 3.2.2 Estimation

We note that the primary mechanism for testing differences between arms does not itself deliver a point estimate for the difference in average outcomes between those arms. Estimated impacts will be based on the corresponding ancova or poisson QMLE regression in equation (1) or (2).

### 3.3 Multiple hypotheses

For our primary hypothesis tests, naive $p$-values reported have correct coverage for those with a priori interest in a particular hypothesis. Beyond this, we control the false discovery rate (FDR) in two ways (Haushofer et al., 2017, Egger et al., 2022). First, we report Anderson (2008) sharpened $q$-values that represent the minimum $q$ value at which each primary hypothesis is rejected, pooling all mentee-related and all mentor-related hypotheses in turn. Second, within each outcome domain, we report $q$-values that control the FDR for all reported $p$-values within that domain. In each case, we report both standard $p$-values and the accompanying minimum $q$-values.

### 3.4 Attrition

We address attrition from the estimating sample at 1-month, 3-month, 6-month, and 9-month follow-up surveys as follows:

1. For all primary and secondary outcome estimating samples, we will test for differential attrition by estimating the probability of attrition as a function of treatments and controls, in a linear probability model that replaces the outcome of equation (1) with a binary measure for individual $i$'s inclusion in the estimating sample in post-intervention time period $t$.

[4] Note that we do not re-residualize the outcome distribution for each of these alternative permutations.

2. If a joint ($F$) test for equality of attrition across all arms in the relevant specification is statistically significant at the 5 percent level, then we modify the estimating equation in (1) or (2) to use inverse probability weights (IPWs). To generate IPWs, the probability of retention in each round will be modelled (via lasso) as a function of treatments, candidate baseline covariates, and their interactions.

3. In the case of statistically significant differential attrition (at the 5% level) in any pairwise comparison being used to estimate differences between arms, we will also estimate Lee (Lee, 2009) bounds as a robustness check.

## References


**Anderson, Michael L.**, "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*, 2008, *103* (484), 1481–1495.

**Baseler, Travis, Thomas Ginn, Robert Hakiza, Helidah Ogude, and Olivia Woldemikael**, "Can Aid Change Attitudes Toward Refugees? Experimental Evidence from Uganda," *Working Paper*, 2022.

**Brooks, Wyatt, Kevin Donovan, and Terence R. Johnson**, "Mentors or teachers? microenterprise training in kenya," *American Economic Journal: Applied Economics*, 2018, *10* (4), 196–221.

**de Mel, Suresh, David McKenzie, and Christopher Woodruff**, "Returns to Capital in Microenterprises: Evidence from a Field Experiment*," *The Quarterly Journal of Economics*, 11 2008, *123* (4), 1329–1372.

**Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker**, "General equilibrium effects of cash transfers: experimental evidence from Kenya," *Econometrica*, November 2022, *90* (6), 2603–2643.

**Haushofer, Johannes, Edward Miguel, Paul Niehaus, and Michael Walker**, "GE effects of cash transfers: Pre-analysis plan for household welfare analysis," American Economic Association RCT Registry, Trial No. AEARCTR-0000505 September 2017.

**Lee, David S**, "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.