The Effects of Phone-Based Surveys on Measurement Quality: When and Why Does Modality Matter? : Pre-Analysis Plan

Travis Baseler* Lipeng Chen† Thomas Ginn‡ Maggie McKenna§

November 2022

1 Overview of Project

Applied research relies heavily on data quality. With the rise of randomized controlled trials (RCTs) in recent years (Webber and Prouse, 2018; de Souza Leão and Eyal, 2019), more and more researchers need to collect their own data through surveys instead of using readily available datasets. How to ensure data quality remains a central issue for survey data.

In this project, we plan to study the impact of survey modality on data quality. Specifically, we introduce three distinct treatments in a field experiment on a set of 900 micro-entrepreneurs in Uganda: 1) whether to conduct the survey in person or over the phone, 2) whether to fix enumerator-respondent pairings across survey rounds, and 3) whether to include a trust-building activity prior to the survey. We also cross-cut the three treatments, which can help us understand how these treatments interplay with each other. We will assess how measurement varies for 1) simple, objective questions, such as basic household and business information, 2) complex, objective questions, such as business profit, 3) subjective, sensitive questions, such as social and political attitudes toward controversial topics, and 4) subjective, non-sensitive questions, such as subjective well-being.

Another research question of interest to us is about experimenter demand effects – respondents may change their answers when they know they are being experimented on. To study whether and how survey modality can mitigate these effects, we follow de Quidt et al. (2018) and Mummolo and Peterson (2019) by explicitly telling respondents the

*University of Rochester, travis.baseler@rochester.edu
†University of Rochester, lchen82@ur.rochester.edu
‡Center for Global Development, tginn@cgdev.org
§University of Rochester, mmckenn9@ur.rochester.edu

results we want in a donation allocation task and then investigate whether their answers vary depending on how they are surveyed.

Our project contributes to the literature that studies the impact of survey methods on data quality (e.g., Di Maio and Fiala, 2020; Garlick et al., 2020; Heath et al., 2020) by cross-cutting survey modality, enumerator familiarity, and trust-building exercise in one field experiment, which enables us to directly compare these three methods and analyze their interaction. Also, our project adds to the literature on experimenter demand effects (de Quidt et al., 2018; Mummolo and Peterson, 2019) by proposing potential solutions through our three treatments, which can help us to get closer to respondents' "true" value.

2 Project Design and Sample

Our sample consists of roughly 900 micro-entrepreneurs, made up of both native Ugandans and refugees, who were all enrolled in a separate study on social and political views about refugees, giving us rich panel information on their business characteristics and social and political attitudes. We will randomize three components of the survey, (1) modality, (2) familiarity of the enumerator to the respondent, and (3) a trust-building exercise at the beginning of the survey.

2.1 Interventions

Intervention 1: Modality

We will randomize whether respondents receive a face-to-face or phone survey.

Intervention 2: Enumerator Familiarity

We will also randomize whether respondents are interviewed by an enumerator with whom they are familiar with. An enumerator is defined as familiar to a respondent if he or she interviewed the respondent in the most recent survey from the study which these entrepreneurs were also enrolled in.

Intervention 3: Trust-building

Lastly, we will randomize whether an enumerator reads a script at the beginning of the survey, which is designed to build trust between the respondent and enumerator.

Treatment Groups

Implementing these three interventions will result in eight treatment groups. We will have four groups as follows: (1) In-person survey with a familiar enumerator, (2) In-person survey with an unfamiliar enumerator, (3) Phone survey with a familiar enumerator, (4)

Phone survey with an unfamiliar enumerator. Each of these groups will then be further split into two groups, where one receives the trust-building script, and the other does not, leaving us with 8 total groups.

To assign survey modality, we stratify the randomization based on respondent identity (Ugandans or refugees), baseline business profit (above median or below median), and baseline supporting for refugees (above median or below median).[1] To assign trusting-building and enumerator familiarity status, we stratify the randomization according to the aforementioned three dimensions and respondents' survey modality treatment status that we get from the first step.

2.2 Hypotheses and Outcomes

Our hypotheses are as follows:

1. There will be significant differences in responses generated by the modality of the survey, particularly for questions which are sensitive, and/or complex.
2. Enumerator familiarity will mitigate the differences driven by survey modality.
3. The trust-building exercise will also mitigate the differences driven by the mode of the survey.

Below, we list four categories of questions: 1) simple, objective questions, 2) complex, objective questions, 3) subjective, sensitive questions, and 4) subjective, non-sensitive questions. Within each category, we further divide the questions into different domains. When testing the hypotheses, we will consider and use an index that includes all questions in each category.

2.2.1 Subjective, Sensitive Outcomes

We believe these hypotheses will hold for **subjective, sensitive outcomes**, such as political and social attitudes on controversial topics.

Domain 1: Support for inclusive refugee hosting.

- Overall, I am in favor of Uganda hosting and assisting refugees. (for Ugandans only)
- Uganda should continue allowing refugees to live outside the settlements, including in Kampala. (for Ugandans only)
- Uganda should continue allowing refugees who live in Uganda right now to work outside the settlements, including in Kampala. (for Ugandans only)

[1] Baseline measures are taken from the latest available wave from a separate study on social and political views on refugees.

- Uganda should continue to provide land to refugees in the settlements. (for Ugandans only)
- Uganda should allow refugees to become full citizens if they have lived in Uganda for a long time and would like to become a Ugandan. As citizens, they would have the right to vote in Ugandan elections.
- Uganda should accept more refugees.

Domain 2: Policy preferences and representation.

- How satisfied are you with the LC1 for this area?
- How satisfied are you with the MP for this area?

Domain 3: Beliefs about economic effects of refugees and non-refugee immigrants (for Ugandans).

- 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?
- Taking everything into consideration, would you say the overall economic effect of foreigners other than refugees on Uganda has been positive, negative, or neutral?
- How about the overall economic effect of foreigners other than refugees on you personally?

Domain 4: Social attitudes toward controversial topics.

- What effect have refugees had on culture in Uganda? (for Ugandans only)
- Social proximity index for Ugandans (measured on 5-point Likert scale; each item is treated as an independent component when calculating the index)
  - I would be comfortable marrying a refugee.
  - I would be comfortable having a refugee marry a member of my family.
  - I would be comfortable having a refugee as a close, personal friend.
  - I would be comfortable having a refugee as a neighbor.
- What effect have foreigners besides refugees had on culture in Uganda? (for Ugandans only)
- Social proximity index for refugees (measured on 5-point Likert scale; each item is treated as an independent component when calculating the index)

  - I would be comfortable marrying a Ugandan.
  - I would be comfortable having a Ugandan marry a member of my family.
  - I would be comfortable having a Ugandan as a close, personal friend.
  - I would be comfortable having a Ugandan as a neighbor.

- A husband is justified in beating his wife if she neglects the children.
- A girl under 18 may refuse marriage proposal even if her parents want her to accept it.
- People with HIV/AIDS should be allowed to hold the same jobs and go to the same schools as all other people.
- I would not allow anyone under age 15, including my child, to work at my business.
- It's okay for any child under the age of 17 to work outside of the home if their family needs the income.

Domain 5: Psychological integration of refugees (for refugees).

- How connected do you feel with Uganda?
- How often do you feel like an outsider in the Uganda?

2.2.2 Complex, Objective Outcomes

We believe expect treatment effects on **complex, objective questions**.

Domain 6: Complex business information.

- Business capital. We sum two variables.
  - If you were to sell all the business-related equipment you own right now, how much do you think you could make?
  - If you were to sell all the inventory you own right now, how much do you think you could make?
- Over the past 7 days, how many hours did you work at this business?
- Over the last 7 days, how much did you pay the members of your household in total?
- In the past 7 days, how many total hours did members work without pay?
- Over the last 7 days, how much did you pay employees from outside of your household in total?

- How much revenue did this business earn in the last 30 days?
- What were the profits of your business during the last 30 days?
- Over the last 30 days, how much money did you invest into upgrading your business?
- How much total business-related debt do you currently have?

Domain 7: Total household income.

- This will be computed as the sum of 4 measures of income.
    - 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 income did you earn in the last 30 days?
    - How much wage income did [other members of your household] earn in the last 30 days?

2.2.3 Subjective, Non-sensitive Outcomes

We think our hypotheses matter less for **subjective, non-sensitive questions.**

Domain 8: Subjective 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 9: Social attitudes toward general topics.

- Generally speaking, most people can be trusted.
- When doing a business deal with another person, most people would try to be fair and not take advantage of you if they got the chance.
- Most of the time, people try to be helpful, and are not just looking out for themselves.

Domain 10: Knowledge of refugees and hosting policy (for Ugandans).

- Are refugees allowed to live outside of the camps or settlements? ("yes" is correct)
- Are any of the international donations to refugees in Uganda shared with Ugandans? ("yes" is correct)

2.2.4 Simple, Objective Outcomes

We do not expect to detect any treatment effects on **simple, objective questions.**

Domain 11: Basic family information.

- How many adults (18 years and older), including yourself, live in your household? Include anyone who usually sleeps under the same roof as you and eats from the same pot.
- How many children (under 18 years old) live in your household?
- How many of those children are 5 years or younger?

Domain 12: Basic business information.

- Do you currently operate a salon/tailor business?
- Is your business officially registered?
- How much do you pay in monthly rent?
- Is this business operated out of the respondent's home?
- How many people from your household, other than you, currently work at this business?
- How many paid employees from outside of your household currently work in this business?
- In the past year, how many times did you take out a loan for your business?

Domain 13: Business practices.

- In the past three months, have you visited at least one of your competitor's businesses to see what prices your competitors are charging?
- In the past three months, have you attracted customers with a special offer on price?
- In the past three months, have you compared the prices or quality offered by alternate suppliers to your business' current suppliers?
- In the past three months, have you run out of stock or raw materials once per month or more?
- In the past three months, have you recorded every purchase and sale made by your business?

- In the past three months, have you 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?
- Over the last 30 days, 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?
- Over the last 30 days, 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?

Domain 14: Social integration of refugees (for refugees).

- In the last 12 months, how often did you eat dinner with Ugandans who are not part of your family?
- Please think about the Ugandans in your neighborhood or your phone contacts. With how many of them did you have a conversation - either by phone, messenger chat, or text exchange - in the last 4 weeks?
- People sometimes participate in different kinds of groups or associations. Do you participate in a group related to your job?

2.2.5 Experimenter Demand Effects

In addition to the randomization on survey modality, we also randomize the sample into two groups for a donation allocation task, in which respondents are asked to divide 3,000 UGX between a refugee program and a Ugandan program. This exercise is designed to measure and bound experimenter demand effects, following de Quidt et al. (2018) and Mummolo and Peterson (2019): some respondents are presented with the statement that "we are hoping that people who hear these instructions and who have received mentoring services/a grant/information/a mentee and payment from YARID, the organization led by refugees, will give more to the *refugee program*" or "we are hoping that people in this study, which is about promoting refugee economic inclusion, and who hear these instructions will give more to the *refugee program*",[2] while others "we are hoping that people in this study, which is about supporting Ugandan small businesses, and who hear these instructions will give more to the *Ugandan program*."

[2]The actual treatment status in a separate study in collaboration with YARID determines what text is shown to the respondents. For example, if the respondents received cash from YARID in the previous study, they are presented with "who have received a grant from YARID" in the priming statement; if the respondents were in the control group, they are presented with the second statement.

We will test whether experimenter demand effects – defined as the impact of priming on the donation task – vary with each of our treatments. Please refer to Section 3 for regression specifications.

3 Statistical Tests

We will measure the effects of survey modality using five different ANCOVA specifications:

$$y_i = \beta P_i + \gamma y_{i0} + \delta M_{i0} + \eta X_i + \alpha_i + \epsilon_i \tag{1}$$

$$y_i = \beta E_i + \gamma y_{i0} + \delta M_{i0} + \eta X_i + \alpha_i + \epsilon_i \tag{2}$$

$$y_i = \beta T_i + \gamma y_{i0} + \delta M_{i0} + \eta X_i + \alpha_i + \epsilon_i \tag{3}$$

$$y_i = \beta_1 P_i + \beta_2 T_i + \beta_3 P_i T_i + \gamma y_{i0} + \delta M_{i0} + \eta X_i + \alpha_i + \epsilon_i \tag{4}$$

$$y_i = \beta_1 P_i + \beta_2 E_i + \beta_3 P_i E_i + \gamma y_{i0} + \delta M_{i0} + \eta X_i + \alpha_i + \epsilon_i \tag{5}$$

y_i is an outcome for individual i, y_{i0} denotes the corresponding baseline outcome (taken from the latest available data from previous baselines), M_{i0} is an indicator for a missing value of y_{i0}, P_i is an indicator for phone vs. face-to-face survey, T_i is an indicator for trusting-building exercise, E_i is an indicator for having a familiar enumerator. X_i is a vector of baseline controls chosen through double LASSO,[3] α_i is a strata fixed effect, and ϵ_i 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 all possible controls from the baseline in each.

In specification (1), we only include the phone vs. face-to-face survey indicator P_i by pooling across fixed enumerator-respondent pairings treatment E_i and trust-building exercise treatment T_i. In specification (2), we pool across P_i and T_i and focus only on the impact of E_i on survey outcomes. Likewise, in specification (3), we focus only on T_i by pooling across P_i and E_i. In specifications (4) and (5), we pool across E_i and T_i, respectively. Both strategies T_i and E_i are commonly implemented by survey firms, and therefore the average treatment effect across these practices is an object of our interest, which also increases power (Muralidharan et al., 2019).

To study how experimenter demand effects vary with treatment status, we will consider interacting our treatment variables with the priming assignment in specifications (1) to (3).

3.1 Measurement

We will transform variables in the following ways:

- Likert scales and other categorical variables will be transformed into binary measures (i.e. "very positive" and "somewhat positive" = 1). Neutrals will be resolved

[3] These controls will be chosen from all baseline variables.

towards the smaller group in order to maximize statistical power; we will report if changing this assignment mechanism changes results. "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 mean.

- Monetary measures including profits, revenues, expenses, capital, and debt:
  - Values will be winsorized at the 1st and 99th percentiles within each survey round.
  - 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.

3.2 Multiple Hypotheses

First, we will report all of the outcomes listed here that are asked in the survey. For the four category summary indices, we will use Westfall-Young step down adjusted p-values to control for the family-wise error rate. Within each domain, we will follow Anderson (2008) to create indices to reduce measurement error and the number of tests, and use sharpened q-values for individual survey questions within domains to control for the false discovery rate.

3.3 Attrition

We will test for differential attrition by P_i along each domain studied in Baseler et al. (2021).

References

Anderson, M. L. (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(484):1481–1495.

Baseler, T., Ginn, T., Ogude, H., and Woldemikael, O. (2021). Can Development Aid Change Attitudes Toward Refugees? Experimental Evidence from Urban Microentrepreneurs in Uganda: Pre-Analysis Plan. *AEA RCT Registry*.

de Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and Bounding Experimenter Demand. *American Economic Review*, 108(11):3266–3302.

de Souza Leão, L. and Eyal, G. (2019). The Rise of Randomized Controlled Trials (RCTs) in International Development in Historical Perspective. *Theory and Society*, 48(3):383–418.

Di Maio, M. and Fiala, N. (2020). Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect. *The World Bank Economic Review*, 34(3):654–669.

Garlick, R., Orkin, K., and Quinn, S. (2020). Call Me Maybe: Experimental Evidence on Frequency and Medium Effects in Microenterprise Surveys. *The World Bank Economic Review*, 34(2):418–443.

Heath, R., Mansuri, G., Rijkers, B., Seitz, W. H., and Sharma, D. (2020). Measuring Employment : Experimental Evidence from Urban Ghana. Technical Report 9263, The World Bank. Publication Title: Policy Research Working Paper Series.

Mummolo, J. and Peterson, E. (2019). Demand Effects in Survey Experiments: An Empirical Assessment. *American Political Science Review*, 113(2):517–529. Publisher: Cambridge University Press.

Muralidharan, K., Romero, M., and Wüthrich, K. (2019). Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments.

Webber, S. and Prouse, C. (2018). The New Gold Standard: The Rise of Randomized Control Trials and Experimental Development. *Economic Geography*, 94(2):166–187.