CENTER FOR GLOBAL DEVELOPMENT

Familiar Faces, Honest Answers? The Impacts of Enumerator Familiarity and Modality on Survey Responses

Travis Baseler, Lipeng Chen, Thomas Ginn

Abstract

Survey data can be distorted through mistrust, discomfort, or demand effects on the part of respondents. We test whether familiar enumerators—those with whom a respondent has completed a prior survey—influence data quality in a panel survey with small business owners in Uganda. We randomly assign respondents to a familiar or a new enumerator and cross-cut a second randomization to an in-person or phone-based interview modality. Across a broad set of outcome types, we observe few impacts of either survey method on means, distributions, or attrition. However, respondents are significantly more likely to give socially desirable answers to new surveyors compared to familiar ones. The effect of familiarity does not interact significantly with a prior experiment conducted with the same sample, suggesting that familiarity influences estimates of levels but not of treatment impacts. These findings suggest that surveys by familiar enumerators can improve the measurement of sensitive beliefs.

KEYWORDS
survey design, enumerator effects, social desirability bias

JEL CODES
C42, C81

Familiar Faces, Honest Answers? The Impacts of Enumerator Familiarity and Modality on Survey Responses

Travis Baseler
University of Rochester, BREAD, and J-PAL
travis.baseler@rochester.edu

Lipeng Chen
Analysis Group
lipengchen@yahoo.com

Thomas Ginn
Center for Global Development
tginn@cgdev.org

Travis Baseler, Lipeng Chen, Thomas Ginn. 2026. "Familiar Faces, Honest Answers? The Impacts of Enumerator Familiarity and Modality on Survey Responses." CGD Working Paper 754. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/familiar-faces-honest-answers-impacts-enumerator-familiarity-and-modality-survey

We are grateful to the International Research Consortium for overseeing data collection, to Innovation for Poverty Action's Peace & Recovery Initiative for funding, to Douglas Blalock and Margaret McKenna for outstanding research assistance, and to the editor and two anonymous referees for helpful comments. This study was approved by the Institutional Review Boards at Harvard (IRB109), Rochester (#4098), the Uganda National Council for Science and Technology (SS 5014), and Mildmay Uganda (0504-2019). This study was pre-registered in the AEA RCT Registry, and the unique identifying number is 10053 (Baseler et al., 2022).

CENTER FOR GLOBAL DEVELOPMENT

2055 L Street, NW Fifth Floor

Washington, DC 20036

202.416.4000

1 Abbey Gardens

Great College Street

London

SW1P 3SE

www.cgdev.org

1 Introduction

Survey data quality is a first-order concern in applied research. A growing literature demonstrates that seemingly minor survey-design choices—from question wording to interview modality—can have large effects on measured outcomes (Beegle et al., 2012, Hsu and McFall, 2015, Garlick et al., 2020). Characteristics of enumerators are known to influence responses, especially on attitudinal or politically sensitive topics (Di Maio and Fiala, 2020). However, little is known about how to leverage enumerator effects to improve data quality. At the same time, the increasing popularity of phone surveying since COVID-19 has raised questions about data comparability across survey modes. Existing evidence finds that moving surveys from in-person to phone does not often change average responses for microenterprise outcomes but can affect variance (Garlick et al., 2020). However, the impact of phone surveying on the measurement of more sensitive outcomes—and whether enumerator effects manifest differently in phone surveys—is largely unknown.

We design and implement a randomized survey experiment in Uganda to test two measurement protocols potentially affecting data quality: 1) enumerator familiarity—pairing respondents either with a familiar enumerator (the same interviewer who surveyed them in a prior wave) or a new enumerator; and 2) survey modality—conducting the interview in person versus by phone.[1] Repeated interactions with an enumerator could establish trust and rapport, leading to better understanding of complicated questions or increased willingness to disclose personal information and views. On the other hand, familiarity could introduce complacency or confirmation bias among enumerators or increase respondents' social image concerns, yielding more socially desirable responses. We cross-cut these two protocols among roughly 900 micro-entrepreneurs who had participated in an earlier study. This factorial design allows us to estimate each protocol's average impact as well as potential interaction effects. Our questionnaire includes a range of outcome measures varying in sensitivity and complexity: simple factual items (e.g., number of employees), complex quantitative items (e.g., profits, revenues), sensitive subjective items (e.g., attitudes toward controversial political or social issues), and non-sensitive subjective items (e.g., life satisfaction).

We find few significant effects of enumerator familiarity or survey modality across the measured means and distributions of these outcomes. However, we find that respondents surveyed by a familiar interviewer (as opposed to a new one) were 0.15 standard deviations

[1] We also cross-randomized a trust-building exercise that attempted to build rapport at the beginning of the survey. However, this exercise was largely unstructured, and audit data revealed that many surveyors spent little time on it. This paper therefore focuses on the other two protocols, while adjusting for multiple testing inclusive of the trust-building assignment per our pre-analysis plan (we describe the trust-building exercise—which yields null impact estimates—further in Appendix A).

less likely to give socially desirable responses on culturally or politically sensitive questions, a difference that survives corrections for multiple hypothesis testing (unadjusted p-value = 0.03; Romano-Wolf p-value = 0.09) and does not differ significantly for phone versus in-person modes. We argue that this effect most likely reflects an increase in truthful reporting, under the assumption that misreporting is biased toward socially desirable responses on average.[2]

Many economic studies are primarily interested in estimating treatment effects rather than means *per se*; if the enumerator familiarity effect does not interact with the treatment being analyzed, then differencing mean responses in the treatment and control groups will net it out even if the effect on mean responses is large. To test whether enumerator familiarity biases treatment effect estimation, we interact it with a prior multi-arm experiment conducted with our study sample. We find no significant differences in treatment impact estimates between new and familiar enumerators, suggesting that the familiarity effect appears as a level difference rather than systematically altering treatment impact estimates.

Neither enumerator familiarity nor survey modality influences average survey response rates, which are statistically indistinguishable at around 78% in all groups. However, heterogeneous response rates across subgroups can introduce attrition bias even when response rates are similar on average. To test for attrition bias beyond impacts on mean response rates, we exploit pre-randomization data collected in a baseline round and characterize the nature of attrition across survey protocols. Reassuringly, we find no significantly differential response rates across a large set of pre-survey characteristics.

Our findings have important implications for data collection methodology. Respondents appear more willing to report socially stigmatized views when they have an established relationship with the enumerator. We believe this reflects an intuitive yet under-explored mechanism: trust and rapport can reduce respondents' fear of judgment or repercussion, increasing the honesty of their responses. Compared to other methods intended to reduce social desirability bias—such as list experiments or randomized-response methods, which typically require complex questionnaire design, reduce statistical power, and can fail internal-validity tests (Blair et al., 2020, Chuang et al., 2021)—assigning respondents to enumerators they have met before is a naturalistic method to increase respondent comfort. Our results also bolster the extant evidence that phone surveys can be a viable substitute for in-person surveys for many types of outcomes, at least in studies like ours where prior in-person contact has been established.

[2]See Section 2.4 for the definition of our social desirability measure and Section 4 for a discussion of its application in other contexts. In our setting, the responses we code as socially desirable were the majority reported view in our sample, the majority reported second-order belief about friends' views (where available), the public position of the Ugandan government, and the public position of international aid organizations at the time of our survey.

Related Literature. Our study contributes to several strands of literature on survey methodology and data quality. A closely related strand concerns enumerator effects. By randomly assigning interviewers to respondents in Uganda, Di Maio and Fiala (2020) demonstrate that enumerator effects are negligible for many standard demographic questions, but can be large for sensitive questions about political opinions. Other studies have similarly found that interviewer attributes such as gender (Flores-Macias and Lawson, 2008) or ethnicity (Adida et al., 2016) can sway responses on topics like gender norms or social trust, though Kadam et al. (2025) find few differences between male and female enumerators on questions about female empowerment in Ethiopia. Our experiment builds on this literature by isolating another specific enumerator attribute: familiarity with the respondent from past survey rounds. While enumerator familiarity has been found to improve survey response rates in some settings (Nicoletti and Buck, 2004, Watson and Wooden, 2009), to our knowledge no study has causally identified its impact on measured outcomes. Our contribution is to show that familiarity can indeed shift data for sensitive attitudinal outcomes.

A second strand of literature examines how survey mode (in person, phone, web, etc.) affects measurement.[3] Many studies report encouragingly small differences in core economic variables across modes, though each mode has pros and cons. In a high-frequency microenterprise panel, Garlick et al. (2020) show that phone and in-person surveys yield similar distributions and means for most outcomes, though phone interviews introduced higher time-series variability and possibly greater sensitivity to social desirability for some questions. Anderson et al. (2024) compare in-person and phone surveys in rural India and find that phone responses are higher on average and more variable for agricultural production data. Despite level differences, treatment-effect estimates were robust across modes (though less precisely estimated with phones). Kadam et al. (2025) find few differences between phone and in-person modes. On the other hand, certain complex metrics are harder to capture by phone, possibly because of greater respondent fatigue: an experiment in urban Ethiopia finds that phone-based consumption modules yield lower consumption estimates (Abate et al., 2023). Sensitive questions can also respond to survey mode: Smarrelli et al. (2026) find audio computer-assisted self-interviews lead to an increase in reporting violence in schools in Malawi relative to face-to-face surveys. Our contribution is to show that differences between phone and in-person surveys are small not only for microenterprise outcomes, but also for

[3] A related strand of research tests how other survey features—such as length, question wording, frequency, or electronic verification—affect data quality. Ambler et al. (2021) find evidence of respondent fatigue in reporting labor activities. de Mel et al. (2009) find that asking firm owners to report profit produces more reliable estimates compared to subtracting reported costs from reported revenues. Fafchamps et al. (2012) find that electronic surveys with built-in consistency checks reduce noise in panel data on microenterprise earnings. See De Weerdt et al. (2020) for a review of experiments on survey methods.

a set of subjective well-being and sensitive political attitude questions. By relying on an experimental sample surveyed in prior rounds, we are also able to test whether attempted contact through phone surveys generates systematically different attrition bias, and find that it does not.

Finally, our findings relate to studies on methods for sensitive question measurement. Recognizing that respondents may under-report taboo or illicit behaviors, survey methodologists have developed indirect techniques such as the list experiment (or item count technique) and the randomized response technique to better measure these behaviors. In theory, these methods provide greater anonymity for individual answers. However, recent evidence has cast doubt on their effectiveness. Chuang et al. (2021), for example, applied internal consistency tests to list experiments and the randomized response technique in sensitive surveys about sexual behavior, and found that the results often failed basic consistency checks. Boudreau et al. (2023) find that another indirect technique—"hard garbling," which entails setting a random subset of responses to the sensitive answer—significantly increases reporting of workplace harassment in Bangladesh. Respondents, however, must still understand and trust the technique as well as its implementation. Many indirect methods also entail a potentially substantial loss of precision because the sensitive response must be recovered from noisy aggregate responses rather than observed individually. For example, the meta-analysis of Blair et al. (2020) calculates that a standard list experiment requires roughly 14 times the sample size to achieve the same precision offered by direct responses. Our findings suggest that enumerator familiarity can increase reporting of less socially desirable outcomes without relying on complex survey design methods that anonymize individual answers. While our approach is limited to panel surveys and introduces some logistical constraints, it has the advantage of being transparent and easily understood by respondents. It can also be used in combination with indirect methods, which may be especially attractive for methods—such as hard garbling—which depend on trust in the enumerator.

2 Study Design

This section describes our study design, including sampling, our two survey protocols, randomization, measurement, and estimation and hypothesis testing. Our main analysis, including outcome measures and estimating equations, was pre-specified at the AEA RCT Registry here.

2.1 Sample

Our study took place in Kampala, the capital city of Uganda. We drew a sample of 1,067 micro-entrepreneurs who had completed recent survey activities from a separate study on views toward refugee policies in Uganda (Baseler et al., 2025). We rely on these pre-randomization (baseline) data both to characterize how attrition depends on survey protocols and to improve precision in impact estimation. Our sample is made up of both native Ugandans (88% of the sample) and refugees operating tailoring shops or salons within 10 kilometers of the city center. At baseline, their businesses made 21 USD in monthly profits and had 550 USD worth of capital stock on average.[4]

Surveys began in November 2022, about eight months after the final survey wave of the Baseler et al. (2025) study, and were completed in March 2023. Overall, we successfully surveyed 836 micro-entrepreneurs from the initial sample of 1,067, giving an average response rate of 78%. Attrition, which is an outcome of interest in this study, is discussed in Section 3.4.

2.2 Survey Protocols

We study two survey protocols potentially affecting data quality:

1. **Enumerator Familiarity.** In the *Same Surveyor* group, we assigned surveyors to the same respondents they had surveyed in a prior wave of the Baseler et al. (2025) study.[5] In the *New Surveyor* group, we assigned surveyors to new respondents.[6]

2. **Modality (In-Person or Phone Survey).** In the *Phone Survey* group, respondents were contacted using phone numbers collected in prior survey activities as part of the Baseler et al. (2025) study. If respondents were successfully reached, the surveyor completed a questionnaire over the phone. In the *In-Person Survey* group, respondents were approached in person at their place of business, which was also listed in prior survey waves. If respondents were successfully found, an identical questionnaire was

[4]Seventy-four percent of our sample took part in the experiments studied in Baseler et al. (2025), where additional summary statistics are presented.

[5]Eighty-seven percent of respondents in the Same Surveyor group were matched to the enumerator who surveyed them in the most recent prior survey wave (March 2022, in person with a mean duration of 52 minutes). The remaining respondents were matched to an enumerator from the May 2021 (in person) or August 2021 (phone) waves, either because they were not found during the March 2022 wave or because the enumerator from the March 2022 wave was no longer available.

[6]Fourteen enumerators were assigned respondents in both the *Same Surveyor* and *New Surveyor* groups. Five enumerators were new to the project and only assigned to the *New Surveyor* group. Results are consistent when enumerator fixed effects are included. Note that all enumerators were of Ugandan nationality.

completed using an in-person interview. In both groups we followed standard data collection protocols, including three attempts (phone calls or in-person visits) on different days of the week and times of the day before recording a respondent as attrited.

2.3 Randomization

We cross-cut assignment to the above two protocols in a sample of 1,067 business owners. Assignment of the modality protocol was stratified on respondent nationality (Ugandans or refugees), baseline business profit (above median or below median), and baseline support for refugee integration (above median or below median). Assignment to the familiarity protocol was stratified along the same dimensions plus assignment of the randomized survey modality.

2.4 Measurement

Our questionnaire was based on earlier versions from the Baseler et al. (2025) study, which included questions on political and social views toward refugees as well as more general demographic, household, and business questions. For the purposes of this study, we organize our questionnaire measures into four category indices based on their expected sensitivity to survey protocol. We hypothesized that sensitive questions (such as stated alignment with prevailing norms) and/or complex questions (such as business profits) would be most responsive. Indices are computed using the method of Anderson (2008).[7] Treatment impacts on each component measure are presented in Appendix B; detailed descriptions of components can be found in our pre-analysis plan here. The four categories are:

1. **Business Formality (Objective and Simple).** This category includes simple business questions (such as number of employees and monthly rent) and business practices (binary indicators for whether the respondent has recently implemented a variety of management practices, such as visiting a competitor to see their prices).

2. **Income and Business Size (Objective and Complex).** This category includes complex business information—capital, hours worked, labor and non-labor costs, revenue, investment, and debt—and non-business wage income.

[7] We implement this method using Stata's *swindex* command, normalizing for weighting using observations assigned to phone surveys by new enumerators. Because the Alignment With Prevailing Norms category includes domains defined only for specific nationalities, we compute this index separately by nationality. Note that our estimating equation includes a nationality control inside the randomization-stratum fixed effect.

3. **Well-Being and Pro-Social Beliefs (Subjective and Non-Sensitive).** This category includes questions on subjective well-being and general social attitudes such as whether most people are trustworthy, fair, and helpful.

4. **Alignment With Prevailing Norms (Subjective and Sensitive).** This category includes measures of support for inclusive refugee hosting and economic and cultural beliefs about refugees, satisfaction with political representatives, and attitudes toward other controversial topics such as domestic violence, HIV/AIDS, and child labor. Note that "true" beliefs are unobservable. Interpreting changes in the measurement of these outcomes requires us to take a stand on the direction of misreporting. We recode all components so that positive responses correspond to agreement with prevailing norms.[8] Under the assumption that respondents do not misreport their views *away* from these prevailing norms, reductions in this measure can be interpreted as closer to the "true" views.

The components of these categories are a subset of our full set of pre-specified measures; measures dropped from our pre-specified index are reported separately in Appendix B.[9] We transform Likert scales and other categorical variables into binary measures, resolving neutrals toward the smaller group. Answers of "Don't know" are treated as missing, and we do not impute missing values for outcome variables. We do impute missing values for control variables using the mean. Monetary measures are winsorized at the 1% level within survey round and deflated using the CPI from the Uganda Bureau of Statistics (UBOS, 2021, 2023).

2.5 Estimation and Hypothesis Testing

We assess the impacts of assignment to each survey protocol using regressions of the form

$$y_i = \beta^X P_i^X + \gamma y_{i0} + \delta M_{i0} + Z_i \eta + \alpha_i + \epsilon_i, \tag{1}$$

where y_i is a survey measure of interest for respondent i, with y_{i0} corresponding to a baseline (pre-experimental) value of y_i; P_i^X is a binary survey protocol assignment indicator for one of

[8] At the time of our study, the Ugandan government was publicly supportive of refugees' living in and moving freely through Uganda, and even opposition groups did not systematically rally against pro-refugee policies (Zhou et al., 2025). The measures in this category are also the majority position in our sample.

[9] We deviate from the pre-analysis plan in order to construct overall index measures that correspond to meaningful economic concepts like business formality or well-being. We therefore exclude family size and social integration from the business formality (objective and simple) index, and knowledge of refugee hosting policy from the well-being and pro-social beliefs (subjective and non-sensitive) index. These changes do not affect the results on enumerator familiarity. The in-person coefficients on these two pre-specified indices are statistically significant.

two protocols $X \in \{\text{Familiar Surveyor}, \text{In-Person Survey}\}$; M_{i0} is an indicator for a missing value of y_{i0}; Z_i is a vector of baseline controls chosen through double-lasso regression from the full set of available baseline variables; α_i is a randomization-stratum fixed effect; and ϵ_i is an error term. We adjust for non-response weights estimated through lasso logit regression using the stratum, treatment, and baseline outcome variables. The coefficient β^X estimates the conditional effect of assignment to protocol X on the average value of y_i, pooling across the omitted protocol assignment. To test for distributional differences, we report p-values from Kolmogorov-Smirnov tests, residualizing y_i according to Equation 1.

Multiple Hypothesis Adjustments. Appendix B reports all outcomes listed in our pre-analysis plan. Those outcomes are organized into 14 conceptual domains; within each domain, we create Anderson (2008) summary indices and report sharpened q-values for domain components to control the false discovery rate. For the four major categories described in Section 2.4, we report step-down adjusted p-values strongly controlling the family-wise error rate by treatment arm.[10]

3 Results

This section presents the impacts of our two randomized survey protocols on our outcomes of interest.

3.1 Measurement Impacts: Enumerator Familiarity

The top panel of Table 1 presents impacts of assignment to a familiar surveyor—one who had surveyed the respondent, most commonly in person, prior to this survey wave—compared to a new surveyor. For both objective and subjective non-sensitive categories, differences are small (magnitudes of 0.07 standard deviations or less) and not statistically significant. However, we observe lower alignment with prevailing norms—our subjective, sensitive category—for respondents assigned to a familiar surveyor. The difference is moderate at 0.15 standard deviations (sd; naïve $p = 0.03$) and survives a family-wise error-rate adjustment with a Romano-Wolf p-value of 0.09.[11] The familiarity effect is somewhat amplified in in-person

[10] These p-values control the probability of committing any Type I error among all of the hypotheses tested. We estimate Romano and Wolf p-values using the Stata command *rwolf2* (Clarke et al., 2020). This package offers the same flexibility as the *wyoung* one we had pre-specified but with faster computation time. We include a pre-specified trust-script treatment in specifications when adjusting for multiple hypothesis testing (see Appendix Table A.1).

[11] This average effect is within the range of enumerator-characteristic effects documented in Di Maio and Fiala (2020). It is larger than the null effects of enumerator characteristics on non-sensitive outcomes but

surveys compared to phone surveys, as shown in Appendix Table B.1, but the difference is not statistically significant. Consistent with a trust mechanism, respondents assigned to familiar enumerators were 9 pp. more likely to say that people can generally be trusted (category-wise FWER $p = 0.05$; see Appendix Table B.9).

Our estimating equations recover conditional mean differences in outcomes, but survey protocols may affect the distribution of outcomes without necessarily changing the mean (for example, by increasing the variance of responses). We test for distributional impacts using Kolmogorov-Smirnov (KS) tests, residualizing each outcome using the same controls as in (1). We find no significant distributional effects of enumerator familiarity, except on alignment with prevailing norms (KS p-value = 0.03).

To unpack our index-based measure, Appendix Tables B.11, B.12, B.13, B.14, and B.15 present impacts on each component of our index capturing alignment with prevailing norms. Respondents were less likely to tell familiar surveyors that they support refugee integration (coefficient = -0.12 sd on domain summary index, category-wise FWER $p = 0.36$). They were also less likely to agree with prevailing norms on other controversial topics (coefficient = -0.15 sd on domain summary index, category-wise FWER $p = 0.12$), including that men should not beat their wives (coefficient = -0.06, category-wise FWER $p = 0.09$). These respondents were also less likely to report that refugees benefit Uganda economically, though the difference is not statistically significant (coefficient = -0.06, category-wise FWER $p = 0.66$). We note that differences for familiar surveyors are not observed for every sensitive outcome, such as reported satisfaction with politicians.

The starkest difference between familiar and new surveyors appears for questions assessing the psychological integration of refugees in Uganda. Refugees were much less likely to report to familiar surveyors that they feel connected with Uganda (coefficient = -27 percentage points on a base of 64%, category-wise FWER $p = 0.01$) and that they don't feel like an outsider in Uganda (coefficient = -21 percentage points on a base of 51%, category-wise FWER $p = 0.09$). Appendix Table B.15 shows that these differences are roughly similar for phone and in-person surveys.

3.2 Measurement Impacts: Survey Modality

The lower panel of Table 1 presents impacts of the in-person survey modality on each of our four category summary index measures. Overall, differences are small. The average

much smaller than the 0.40–0.55 sd impacts of urban enumerators on highly sensitive outcomes in a rural population like support for opposition parties. Decomposing our index-based effect, we find large impacts of 0.42–0.56 sd on whether refugees report that they feel connected to Uganda, or feel like an outsider in Uganda, and smaller impacts on outcomes that are less sensitive, like expressed support for refugee hosting (0.12 sd).

difference in our summary index is within 0.11 standard deviations (sd) for business formality (based on objective, simple measures), income and business size (based on objective, complex measures), well-being and pro-social beliefs (based on subjective, non-sensitive measures) and alignment with prevailing norms (based on subjective, sensitive measures). No FWER-adjusted p-value is below or near 0.10.

However, there are modest negative impacts of in-person surveys on measured business formality (objective, simple) and well-being and pro-social beliefs (subjective, non-sensitive) which are marginally significant without multiple hypothesis corrections (unadjusted p-values = 0.10 and 0.12 respectively). Appendix Table B.2—which presents treatment impacts on each component of the index measures shown in Table 1—shows that the only statistically significant difference appears in whether the business is reported as operational (coefficient = -4 pp. on a base of 97%, within-domain $q = 0.05$). While this magnitude is arguably small and the high FWER-adjusted p-value cautions against over-interpreting this difference, it is possible that in-person surveys—which typically occurred at respondents' place of business—allow enumerators to more easily observe whether the business is operational. Note that this difference is not consistent with strategic under-reporting of business openness to shorten survey length, as such behavior would likely generate greater under-reporting over the phone. Instead, it may reflect mis-classification of marginal businesses, for example if respondents are embarrassed to report that their business has shut down but find it difficult to hide this in person.

As shown in Appendix Table B.9, the modest negative impact of in-person surveys on measured well-being and pro-social beliefs appears to be driven by two outcomes: the respondent's agreement with the claims "people can be trusted" and "people are helpful" (effect sizes = 17% and 9% respectively, within-domain q-values = 0.07). While we remain wary of over-interpreting a handful of differences given the large number of hypothesis tests and the high FWER-adjusted p-value for this category, it is possible that respondents feel slightly more comfortable disagreeing with pro-social views in person, where rapport with the surveyor is presumably greater on average.

We find no significant distributional impacts of survey modality on business formality, income & business size, or subjective, sensitive beliefs, as shown in the middle panel of Table 1. We reject distributional equality for well-being & pro-social beliefs (KS p-value = 0.01). This reflects a distributional shift—rather than a mean-preserving spread—with slightly lower mean well-being as measured in in-person surveys. This may reflect a slight tendency for phone surveys to bias respondents toward socially desirable answers, though the mean difference is not significant after multiple hypothesis testing adjustments ($p = 0.38$).

3.3 Effects on Treatment Impact Estimation

While we find significant impacts of enumerator familiarity on average reported attitudes for sensitive outcomes, it does not follow that (a lack of) enumerator familiarity must distort treatment impact estimation. We would expect biases to appear in treatment impact estimates only if the treatment amplifies or reduces the effect of familiarity.

To test for these biases, we estimate the sensitivity of treatment impact estimates from an earlier multi-arm experiment conducted with our sample. Our sample was randomly exposed to one of five interventions including cash transfers, information about refugee hosting policies, and across- or within-nationality business mentorship programs (Baseler et al., 2025). Three of these—cash, information, and cash combined with information—generated large, long-run impacts on components of our "alignment with prevailing norms" index.

We find few significant interactions between enumerator familiarity or survey modality on these treatment impact estimates, as shown in Table 2. Across ten interaction terms (five interventions interacted with two protocols), a handful of individual coefficients are large in magnitude, and one phone-survey interaction term is significantly different (Romano-Wolf p-value = 0.03). While this difference survives multiple hypothesis testing corrections, it appears in a comparison arm (same-nationality mentorship) that was not directly related to the outcomes we measure and did not affect norm-congruent views on average. Taken together, we find little evidence of systematic interactions between treatment impacts and survey modality or enumerator familiarity.[12]

3.4 Non-Response

We find similar response rates to the survey across our two protocols, as shown in Table 3 column 1. Surveys attempted by phone were slightly more likely to be completed successfully, although the difference is small and statistically insignificant (2 pp. difference, mean response rate = 78%, p-value on difference = 0.34, q-value = 1).

Using survey data from our full experimental sample collected during a past project (Baseler et al., 2025), we characterize how non-response varies by survey protocol by regressing a response indicator on a protocol assignment indicator, ten pre-specified summary indices—covering business measures, household income, subjective well-being, and agreement with potentially sensitive political views—and their interactions with the protocol assignment indicator. Results are shown in Table 3 columns 2–3. We find no evidence of significantly differential attrition along these measures: no q-value on either pre-experimental

[12] We additionally test whether enumerator familiarity affects experimenter demand, as these effects could distort treatment impact estimates (de Quidt et al., 2018, Mummolo and Peterson, 2019). Results, presented in Appendix A, are consistent with a limited interaction between familiarity and demand effects.

summary indices or their treatment interactions is significant at the 10% level. These findings are reassuring for researchers relying on phone surveys. In our setting, response rates are similar for phone and in-person survey attempts, both on average and across several economic and social outcomes.

4 Conclusion

This paper estimates the impacts of two survey protocols on the measurement of a broad set of outcomes. We find that most measures are robust to surveying in person versus over the phone. This finding echoes much of the prior work on modality effects, though some studies find downsides of phone-based surveys, and we note that modality effects may be different at the first point of contact.[13] We also find broadly consistent results when assigning familiar versus new surveyors to respondents. However, one main result survives multiple hypothesis testing adjustments and warrants further investigation in our view: maintaining the same interviewer across survey waves leads respondents to provide answers that are less aligned with prevailing social norms on politically and culturally sensitive questions. The magnitude of this effect—0.15 standard deviations on our sensitive-attitudes index—is large relative to many program impacts.

We believe the increased reporting of minority views to familiar enumerators most likely reflects an increase in truthful reporting on average. This interpretation holds under the assumption that misreporting is biased toward socially desirable responses, which is consistent with other literature on sensitive questions (Boudreau et al., 2023). However, respondents' true underlying views are unknown, and alternative interpretations are possible. For example, respondents may increasingly anchor to enumerators' views across repeated surveys: if respondents perceive those views as norm-*in*congruent, then this behavior could shift responses away from other norms. We find this explanation plausible in general but unlikely in our context, where aid and research organizations and their staff predominantly express views—to the extent they express any personal views in professional settings—in the direction of our norm-alignment measure.

For survey methodology, these results point to enumerator familiarity as a simple alterna-

[13]For example, Abate et al. (2023) find evidence of fatigue effects, and Garlick et al. (2020) find an increase in variance from phone surveying. Our questionnaire was short and relied on summary or recall-based items for business financials rather than itemized consumption recall, likely reducing the scope for fatigue-driven mode effects. We also note that our sample consists of urban micro-entrepreneurs who had already completed at least one prior survey wave and were highly phone-literate. Comfort with remote interviews may have been higher in our sample than in settings where phone surveying is newer or less familiar. Finally, our sample consists of respondents who had previously been interviewed in person. We find it plausible that modality might affect respondent comfort to the greatest extent the first time that respondent is contacted.

tive (when panel designs permit it) to indirect techniques like list experiments or randomized response, which risk confusing respondents and further distorting measurement (Chuang et al., 2021), and as a potential complement to hard garbling (Boudreau et al., 2023). We note that assigning familiar enumerators to respondents may not be feasible or desirable depending on the logistical constraints involved, and that it is not mutually exclusive with indirect methods like list experiments. Researchers seeking to use this method should pay careful attention to which norms are likely to be distorting responses, and in which direction.[14]

Open questions include how to mimic the benefits of surveyor familiarity in cross-sectional surveys—for example, through community-based enumerator assignment or extended introductions—and how far such strategies can go without introducing other biases such as complacency or confirmatory probing. More broadly, our findings reinforce that survey protocols themselves can causally shape what we measure. We believe future work can continue to test design choices that align respondent comfort with researchers' goals of truthful reporting.

[14] A given survey question may be subject to multiple, possibly competing norms: for example local vs. national, community vs. family, or elite vs. popular. Following the substantial literature on experimenter demand effects, our view is that responses are likely to be distorted on average toward the views of the research team—as respondents perceive them. For studies where reliable measures of sensitive beliefs are important, attempting to characterize respondents' views of the study team's goals or beliefs through qualitative work may be helpful.

Table 1: Impacts of Survey Protocol on Four Measurement Indices

Description of Measures:	Business Formality	Income & Business Size	Well-Being & Pro-Social Beliefs	Alignment With Prevailing Norms
Measure Category:	Objective, Simple	Objective, Complex	Subjective, Non-Sensitive	Subjective, Sensitive
Familiar vs. New Surveyor				
Familiar Surveyor	0.04	-0.07	0.03	-0.15
	(0.07)	(0.06)	(0.06)	(0.07)
	[0.55]	[0.28]	[0.61]	[0.03]
	⟨0.79⟩	⟨0.60⟩	⟨0.79⟩	⟨0.09⟩
KS p-Value	0.81	0.63	0.59	0.03
Observations	836	836	836	836
In-Person vs. Phone Survey				
In-Person Survey	-0.11	-0.01	-0.10	-0.03
	(0.07)	(0.06)	(0.07)	(0.07)
	[0.10]	[0.90]	[0.12]	[0.66]
	⟨0.38⟩	⟨0.90⟩	⟨0.38⟩	⟨0.90⟩
KS p-Value	0.21	0.86	0.01	0.74
Observations	836	836	836	836

Each outcome is a summary index with mean 0, standard deviation 1 (see Section 2.4 for details). Heteroskedasticity-robust standard errors in parentheses; p-values in brackets; Romano-Wolf family-wise p-values estimated within rows in angle brackets.

Table 2: Impacts of Surveyor Familiarity and Modality on Treatment Impact Estimation

Intervention:	Same-Nationality Mentorship	Cross-Nationality Mentorship	Cash Grant	Information on Aid-Sharing	Cash + Information
Familiar vs. New Surveyor					
Treatment × Familiar Surveyor	0.32	0.06	-0.09	0.18	-0.14
	(0.27)	(0.25)	(0.19)	(0.21)	(0.19)
	[0.23]	[0.80]	[0.62]	[0.40]	[0.45]
	⟨0.63⟩	⟨0.94⟩	⟨0.94⟩	⟨0.80⟩	⟨0.80⟩
Treatment	-0.36	0.29	0.63	0.12	0.61
	(0.18)	(0.18)	(0.18)	(0.20)	(0.18)
	[0.05]	[0.10]	[0.00]	[0.54]	[0.00]
Observations	600	600	600	600	600
In-Person vs. Phone Survey					
Treatment × In-Person Survey	0.75	-0.26	-0.18	-0.27	-0.06
	(0.26)	(0.24)	(0.20)	(0.21)	(0.19)
	[0.00]	[0.27]	[0.37]	[0.20]	[0.75]
	⟨0.03⟩	⟨0.62⟩	⟨0.62⟩	⟨0.60⟩	⟨0.75⟩
Treatment	-0.65	0.42	0.67	0.34	0.56
	(0.21)	(0.19)	(0.18)	(0.19)	(0.18)
	[0.00]	[0.03]	[0.00]	[0.06]	[0.00]
Observations	600	600	600	600	600

Each outcome is a summary index of alignment with prevailing norms with mean 0, standard deviation 1 (see Section 2.4 for details). Each column shows treatment impact estimates for an intervention from Baseler et al. (2025) interacted with interviewer continuity (upper panel) or modality (lower panel). All regressions are estimated through (1) adding fixed effects for the full set of Baseler et al. (2025) intervention assignments, the randomization strata used to assign them, and the modality (or continuity) assignment (not shown in output). Heteroskedasticity-robust standard errors in parentheses; p-values in brackets; Romano-Wolf family-wise p-values estimated within rows in angle brackets.

Table 3: Response Rates across Protocols and Baseline Data

	Average Response by Treatment	Response Interactions (In-Person Survey)	Response Interactions (Familiar Surveyor)
Treat (In-Person Survey)	-0.02	-0.03	
	(0.03)	(0.02)	
	[1.00]	[1.00]	
Treat (Familiar Surveyor)	0.00		0.02
	(0.03)		(0.02)
	[1.00]		[1.00]
Social Desirability		-0.01	-0.01
		(0.02)	(0.02)
		[1.00]	[1.00]
Household Income		0.02	0.03
		(0.02)	(0.02)
		[1.00]	[1.00]
Subjective Well-Being		0.03	-0.02
		(0.02)	(0.02)
		[0.74]	[1.00]
Business Size		-0.00	0.00
		(0.02)	(0.02)
		[1.00]	[1.00]
Treat × Social Desirability		-0.02	-0.02
		(0.02)	(0.02)
		[1.00]	[1.00]
Treat × Household Income		-0.01	-0.03
		(0.03)	(0.03)
		[1.00]	[1.00]
Treat × Subjective Well-Being		-0.06	0.05
		(0.03)	(0.03)
		[0.34]	[1.00]
Treat × Business Size		0.04	0.03
		(0.02)	(0.03)
		[1.00]	[1.00]
Outcome Mean	0.78	0.78	0.78
Observations	1,067	1,067	1,067

Outcome is an indicator for whether the survey was completed. Regressions in columns 2 and 3 include ten pre-specified summary indices and their interactions with a single treatment assignment indicator (a subset of baseline indicators and interactions is omitted from the table). Heteroskedasticity-robust standard errors in parentheses; sharpened q-values (Anderson, 2008) in brackets are estimated within column.

References

Abate, Gashaw T., Alan de Brauw, Kalle Hirvonen, and Abdulazize Wolle, "Measuring consumption over the phone: Evidence from a survey experiment in urban Ethiopia," *Journal of Development Economics*, 2023, *161*, 103026.

Adida, Claire, Karen Ferree, Daniel Posner, and Amanda Robinson, "Who's Asking? Interviewer Coethnicity Effects in African Survey Data," *Comparative Political Studies*, 03 2016, *49*.

Ambler, Kate, Sylvan Herskowitz, and Mywish K. Maredia, "Are we done yet? Response fatigue and rural livelihoods," *Journal of Development Economics*, 2021, *153*, 102736.

Anderson, Ellen, Travis J. Lybbert, Ashish Shenoy, Rupika Singh, and Daniel Stein, "Does survey mode matter? Comparing in-person and phone agricultural surveys in India," *Journal of Development Economics*, 2024, *166*, 103199.

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, Lipeng Chen, Thomas Ginn, and Margaret McKenna, "The Effects of Phone-Based Surveys on Measurement Quality: When and Why Does Modality Matter?," 2022. AEA RCT Registry. June 28. https://doi.org/10.1257/rct.10053-3.0.

_ , Thomas Ginn, Robert Hakiza, Helidah Ogude-Chambert, and Olivia Woldemikael, "Can Redistribution Change Policy Views? Aid and Attitudes toward Refugees," *Journal of Political Economy*, 2025, *133* (9).

Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson, "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania," *Journal of Development Economics*, 2012, *98* (1), 3–18.

Blair, Graeme, Alexander Coppock, and Margaret Moor, "When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments," *American Political Science Review*, 2020, *114* (4), 1297–1315.

Boudreau, Laura E, Sylvain Chassang, Ada Gonzalez-Torres, and Rachel Heath, "Monitoring Harassment in Organizations," Working Paper 31011, National Bureau of Economic Research 2023.

Chuang, Erica, Pascaline Dupas, Elise Huillery, and Juliette Seban, "Sex, lies, and measurement: Consistency tests for indirect response survey methods," *Journal of Development Economics*, 2021, *148*, 102582.

Clarke, Damian, Joseph P Romano, and Michael Wolf, "The Romano–Wolf multiple-hypothesis correction in Stata," *The Stata Journal*, 2020, *20* (4), 812–843.

de Mel, Suresh, David McKenzie, and Christopher Woodruff, "Measuring Microenterprise Profits: Must We Ask How the Sausage Is Made?," *Journal of Development Economics*, 2009, *88* (1), 19–31.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

De Weerdt, Joachim, John Gibson, and Kathleen Beegle, "What Can We Learn from Experimenting with Survey Methods?," *Annual Review of Resource Economics*, 2020, *12*

(1), 431–447.

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

Fafchamps, Marcel, David McKenzie, Simon Quinn, and Christopher Woodruff, "Using PDA Consistency Checks to Increase the Precision of Profits and Sales Measurement in Panels," *Journal of Development Economics*, 2012, *98* (1), 51–57.

Flores-Macias, Francisco and Chappell Lawson, "Effects of Interviewer Gender on Survey Responses: Findings from a Household Survey in Mexico," *International Journal of Public Opinion Research*, 02 2008, *20.*

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

Hsu, Joanne W. and Brooke H. McFall, "Mode Effects in Mixed-Mode Economic Surveys: Insights from a Randomized Experiment," Finance and Economics Discussion Series 2015-008, Board of Governors of the Federal Reserve System 2015.

Kadam, Aditi, Ellen B McCullough, Tamara J McGavock, and Nicholas Magnan, "Who is asking and how? Effects of survey mode and enumerator gender on measuring women's life experience," *World Development*, 2025, *195*, 107078.

Mummolo, Jonathan and Erik Peterson, "Demand Effects in Survey Experiments: An Empirical Assessment," *American Political Science Review*, 2019, *113* (2), 517–529.

Nicoletti, Cheti and Nicholas Buck, "Explaining Interviewee Contact and Co-operation in the British and German Household Panels," *Institute for Social and Economic Research, ISER working papers*, 01 2004.

Smarrelli, Gabriela, Esme Kadzamira, Thi Le, and Tionge Saka, "Advancing the Measurement of Violence in and Around Schools: Evidence from Malawi," Working Paper 751, Center for Global Development 2026.

UBOS, "UGANDA CONSUMER PRICE INDEX: July 2021," Technical Report, Uganda Bureau of Statistics 2021.

_ , "UGANDA CONSUMER PRICE INDEX: December 2023," Technical Report, Uganda Bureau of Statistics 2023.

Watson, Nicole and Mark Wooden, "Identifying Factors Affecting Longitudinal Survey Response," in Peter Lynn, ed., *Methodology of Longitudinal Surveys*, Chichester: John Wiley & Sons, Ltd, 2009, chapter 10, pp. 157–181.

Zhou, Yang-Yang, Naijia Liu, Shuning Ge, and Guy Grossman, "Liberalizing Refugee Hosting Policies without Losing the Vote," Working Paper 2025.

Appendix for "Familiar Faces, Honest Answers? The Impacts of Enumerator Familiarity and Survey Modality on Survey Responses"

A Additional Study Details

A.1 Scripts

Confidentiality script (given to all respondents):

This discussion will remain confidential and your name will not be mentioned in the study. If you choose to participate in this study it is voluntary, and you may decline to answer any question and you may stop the survey at any time.

You are encouraged to be as honest as possible; no opinions will be judged. Your answers will not affect any programs you are currently participating in or could be eligible for in the future — let me repeat that this survey is purely for research purposes.

To thank you for your time, you will be compensated 10,000 UGX for participating.

Trust-building script (given to respondents assigned to trust script):

Since some people may be uncomfortable answering the survey, especially with someone who you don't know well, I want to take a minute to tell you about me, why I think this project is important, and answer any questions you have about me or the work.

My name is _____. I have been working for the International Research Consortium for _____ (years / months). I have worked on (many / a few) other research projects like this one, including studies on _____ and _____. I (studied / am studying) _____ (field) at _____ (university). I have lived in Kampala for _____ years, and I live in _____ (neighborhood) with my _____ (family).

I enjoy this work because I get to meet many different people like you and hear about the issues that are important to them and their communities. I also enjoy this work because I think it's important. The results of our studies, including this one, are discussed with the Ugandan government and international organizations so that programs can better help people in poverty.

What questions do you have about me, the International Research Consortium, or this project? I am happy to answer them.

A.2 Measurement Impacts: Trust-Building Script

The bottom panel of Table A.1 presents impacts of the trust-building script. Differences are small (magnitudes of 0.01 to 0.09 sd) and statistically insignificant across all four category indices. No KS statistic is statistically significant. For our measure capturing alignment with prevailing norms, the impact of the trust script is 0.02 sd (category-wise FWER $p = 0.96$). This result suggests that the effect of familiar surveyors on reported alignment with prevailing views—driven, for example, by greater trust or rapport between the surveyor and respondent—is not easily compensated for with trust-building exercises conducted in a single survey. This result is consistent with Boudreau et al. (2023), who find minimal effects from a similar rapport-building exercise.

One potential explanation for the null impacts of the trust script is that surveyors did not take the activity seriously. Our trust-building script was only partly structured, and data captured during automated survey audits suggest that many surveyors spent little time on it. We view those decisions as somewhat informative: despite spending significant time on other, equally audited parts of the survey, many surveyors apparently viewed the trust-building exercise as not worth much time.

A.3 Experimenter Demand Exercise

To measure experimenter demand effects and assess whether they vary by survey protocol, we used the "strong demand treatment" method of de Quidt et al. (2018). Specifically, we explicitly prime a random subset of respondents with an "expected answer" in a donation-allocation game. Each respondent decided how to allocate 3,000 Ugandan Shillings (≈ 0.84 USD) between a non-profit supporting Ugandans and one supporting refugees in Uganda.[15] Before making their decision, one group received the instructions,

> *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,*

while the other group received the instructions,

> *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.*

We estimate whether experimenter demand effects vary by survey protocol by interacting our protocol assignment indicators with assignment to the randomized demand-elicitation instructions. Our internal audits show that while many surveys appeared to adhere to this protocol, some paraphrased the instructions in a way that could have obscured the demand treatment. Since this will tend to bias our coefficients toward zero, we view these findings as suggestive.

Table A.2 presents impacts of the pro-refugee instruction, and its interaction with survey protocol, on donations to the refugee organization. We find that demand effects appear small

[15] Respondents received 10,000 Ugandan Shillings for participating in the survey.

overall and do not differ significantly by survey protocol. Coefficients are small (magnitudes of 4% of the endowment or less) and are not statistically significant. These results suggest that interactions between demand effects and survey modality or enumerator familiarity are limited.

Table A.1: Impacts of Trust-Building Script

Description of Measures:	Business Formality	Income & Business Size	Well-Being & Pro-Social Beliefs	Alignment With Prevailing Norms
Measure Category:	Objective, Simple	Objective, Complex	Subjective, Non-Sensitive	Subjective, Sensitive
Trust Script vs. No Script				
Trust Script	0.09	-0.02	0.01	0.03
	(0.07)	(0.06)	(0.07)	(0.07)
	⟨0.60⟩	⟨0.96⟩	⟨0.96⟩	⟨0.96⟩
KS p-Value	0.62	1.00	0.96	0.65
Observations	836	836	836	836
Trust and Familiar Surveyor Interactions				
Familiar Surveyor	0.00	0.01	-0.02	-0.15
	(0.10)	(0.08)	(0.09)	(0.10)
	⟨1.00⟩	⟨1.00⟩	⟨1.00⟩	⟨0.72⟩
Trust Script	0.05	0.06	-0.05	0.03
	(0.10)	(0.09)	(0.09)	(0.10)
	⟨1.00⟩	⟨0.99⟩	⟨1.00⟩	⟨1.00⟩
Familiar Surveyor × Trust Script	0.08	-0.16	0.11	0.00
	(0.14)	(0.12)	(0.13)	(0.13)
	⟨1.00⟩	⟨0.83⟩	⟨0.98⟩	⟨1.00⟩
Observations	836	836	836	836
Trust and Modality Interactions				
In-Person Survey	-0.12	-0.05	-0.02	0.06
	(0.10)	(0.09)	(0.09)	(0.10)
	⟨0.87⟩	⟨0.96⟩	⟨0.98⟩	⟨0.96⟩
Trust Script	0.08	-0.06	0.10	0.12
	(0.09)	(0.09)	(0.09)	(0.09)
	⟨0.94⟩	⟨0.96⟩	⟨0.93⟩	⟨0.85⟩
In-Person Survey × Trust Script	0.01	0.09	-0.18	-0.18
	(0.14)	(0.12)	(0.13)	(0.13)
	⟨0.98⟩	⟨0.96⟩	⟨0.83⟩	⟨0.83⟩
Observations	836	836	836	836

Each outcome is a summary index with mean 0, standard deviation 1 (see Section 2.4 for details). Heteroskedasticity-robust standard errors in parentheses; Romano-Wolf (RW) family-wise p-values estimated within rows in angle brackets.

Table A.2: Experimenter demand effects appear similar across survey protocols.

Outcome: Share Donated to Refugees	Average Demand Effect	Demand Interactions (In-Person Survey)	Demand Interactions (Familiar Surveyor)	Demand Interactions (Trust Script)
Pro-Refugee Elicitation	-0.02	-0.00	-0.00	0.00
	(0.01)	(0.02)	(0.02)	(0.02)
	[1.00]	[1.00]	[1.00]	[1.00]
Pro-Refugee Elicitation × Treat		-0.03	-0.04	-0.04
		(0.03)	(0.02)	(0.02)
		[1.00]	[0.75]	[0.75]
In-Person Survey		0.00		
		(0.02)		
		[1.00]		
Familiar Surveyor			0.02	
			(0.01)	
			[1.00]	
Trust Script				-0.01
				(0.01)
				[1.00]
Control Mean	0.47	0.48	0.46	0.47
Observations	836	836	836	836

Outcome is the share donated to a refugee-supporting organization (with the remainder going to a Ugandan-supporting organization) in a dictator game. *Pro-Refugee Elicitation* is an indicator for the randomized assignment to a pre-refugee demand elicitation script following the methodology of de Quidt et al. (2018), with the comparison group receiving a pro-Ugandan elicitation script. Each regression in columns 2–4 includes the pro-refugee demand treatment and its interaction with a single survey protocol assignment indicator. Heteroskedasticity-robust standard errors in parentheses; sharpened q-values (Anderson, 2008) in brackets.

B Components of Category Indices and Other Pre-Specified Results

Modality and Familiarity Interaction Effects

Table B.1: Survey Protocol Interaction Effects

Description of Measures:	Business Formality	Household & Business Finances	Well-Being & Pro-Social Beliefs	Alignment With Prevailing Norms
Measure Category:	Objective, Simple	Objective, Complex	Subjective, Non-Sensitive	Subjective, Sensitive
Surveyor and Modality Interactions				
Familiar Surveyor	0.08	-0.15	0.12	-0.08
	(0.09)	(0.09)	(0.09)	(0.09)
	[0.34]	[0.10]	[0.20]	[0.37]
	⟨0.93⟩	⟨0.61⟩	⟨0.85⟩	⟨0.93⟩
In-Person Survey	-0.07	-0.08	-0.02	0.04
	(0.10)	(0.09)	(0.09)	(0.10)
	[0.49]	[0.35]	[0.86]	[0.70]
	⟨0.93⟩	⟨0.93⟩	⟨0.93⟩	⟨0.93⟩
Familiar Surveyor × In-Person Survey	-0.09	0.15	-0.18	-0.14
	(0.14)	(0.12)	(0.13)	(0.13)
	[0.54]	[0.21]	[0.18]	[0.28]
	⟨0.93⟩	⟨0.85⟩	⟨0.83⟩	⟨0.91⟩
Observations	836	836	836	836

Each outcome is a summary index with mean 0, standard deviation 1 (see Section 2.4 for details). Heteroskedasticity-robust standard errors in parentheses; p-values in brackets; Romano-Wolf family-wise p-values estimated within rows in angle brackets.

Results in the *Business Formality (Objective and Simple)* Category

Table B.2: Basic Business Information

	Operate Business	Business Registered	Monthly Rent (Inv Hyp Sin)	Business Operated from Home	Household Employees	Outside Employees	Business Loans	Summary Index
Familiar vs. New Surveyor								
Familiar Surveyor	0.02	-0.01	-0.02	0.07	0.01	-0.13	0.03	0.06
	(0.02)	(0.03)	(0.04)	(0.03)	(0.04)	(0.06)	(0.07)	(0.07)
	[1.00]	[1.00]	[1.00]	[0.11]	[1.00]	[0.11]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{0.17}	{1.00}	{0.17}	{1.00}	{1.00}
Control Mean	0.94	0.41	44.46	0.72	0.24	0.56	0.72	-0.05
Observations	836	836	780	836	836	835	830	836
In-Person vs. Phone Survey								
In-Person Survey	-0.04	-0.01	-0.00	-0.03	-0.04	0.01	-0.05	-0.12
	(0.01)	(0.03)	(0.04)	(0.03)	(0.04)	(0.06)	(0.07)	(0.07)
	[0.05]	[1.00]	[1.00]	[0.85]	[0.85]	[1.00]	[1.00]	[0.38]
	{0.17}	{1.00}	{1.00}	{0.71}	{0.71}	{1.00}	{0.82}	{0.47}
Control Mean	0.97	0.43	42.05	0.76	0.25	0.48	0.80	0.06
Observations	836	836	780	836	836	835	830	836
Trust Script vs. No Script								
Trust Script	0.00	-0.00	-0.00	-0.05	0.03	0.05	0.08	0.01
	(0.02)	(0.03)	(0.04)	(0.03)	(0.04)	(0.06)	(0.07)	(0.07)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.94	0.42	43.85	0.78	0.23	0.49	0.72	-0.01
Observations	836	836	780	836	836	835	830	836

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Table B.3: Business Practices

	Check Competitor Prices	Give Special Offers	Compare btw Suppliers	Use Up Stock	Record Keeping	Have Written Budget	Marketing	Sell on Credit	Buy on Credit	Summary Index
Familiar vs. New Surveyor										
Familiar Surveyor	-0.03	-0.00	-0.01	-0.00	-0.00	0.02	-0.08	0.11	0.01	-0.01
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.07)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[0.01]	[0.01]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{0.03}	{0.02}	{1.00}	{1.00}
Control Mean	0.78	0.90	0.82	0.51	0.55	0.48	0.22	0.32	0.10	-0.00
Observations	836	836	836	836	836	836	836	836	836	836
In-Person vs. Phone Survey										
In-Person Survey	-0.01	-0.03	-0.01	-0.05	0.00	0.01	0.02	-0.03	-0.00	-0.07
	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.07)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	{1.00}	{0.57}	{1.00}	{0.57}	{1.00}	{1.00}	{0.72}	{0.71}	{1.00}	{0.71}
Control Mean	0.78	0.91	0.82	0.53	0.54	0.47	0.19	0.39	0.10	0.03
Observations	836	836	836	836	836	836	836	836	836	836
Trust Script vs. No Script										
Trust Script	0.01	0.04	0.04	0.02	-0.01	0.03	0.01	0.02	0.03	0.12
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.07)
	[0.92]	[0.55]	[0.55]	[0.92]	[0.98]	[0.87]	[0.92]	[0.92]	[0.55]	[0.55]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.77	0.88	0.80	0.50	0.53	0.48	0.19	0.37	0.08	-0.07
Observations	836	836	836	836	836	836	836	836	836	836

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Measures excluded from our pre-specified index construction

Table B.4: Basic Family Information

	Adults	Children Aged btw 6 and 18	Children Under 5	Summary Index
Familiar vs. New Surveyor				
Familiar Surveyor	0.11	-0.01	-0.07	-0.02
	(0.08)	(0.07)	(0.05)	(0.06)
	[0.44]	[0.87]	[0.44]	[0.87]
	{0.67}	{1.00}	{0.67}	{1.00}
Control Mean	2.65	1.38	0.96	-0.00
Observations	834	836	836	836
In-Person vs. Phone Survey				
In-Person Survey	-0.12	-0.13	-0.04	-0.12
	(0.08)	(0.07)	(0.05)	(0.06)
	[0.18]	[0.16]	[0.26]	[0.16]
	{0.57}	{0.47}	{0.82}	{0.39}
Control Mean	2.75	1.41	0.92	0.02
Observations	834	836	836	836
Trust Script vs. No Script				
Trust Script	-0.01	-0.01	-0.00	-0.01
	(0.08)	(0.07)	(0.05)	(0.06)
	[1.00]	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	2.71	1.35	0.95	0.00
Observations	834	836	836	836

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Table B.5: Social Integration of Refugees (for Refugees)

	Eat Dinner w Ugandans	Converse w Ugandans	Join Work-related Groups	Summary Index
Familiar vs. New Surveyor				
Familiar Surveyor	-0.06	-0.16	0.12	-0.04
	(0.08)	(0.08)	(0.08)	(0.20)
	[0.50]	[0.27]	[0.29]	[0.75]
	{1.00}	{0.31}	{0.67}	{1.00}
Control Mean	0.69	0.57	0.42	0.08
Observations	149	149	149	149
In-Person vs. Phone Survey				
In-Person Survey	-0.14	0.19	0.13	0.31
	(0.08)	(0.08)	(0.08)	(0.19)
	[0.09]	[0.08]	[0.09]	[0.09]
	{0.47}	{0.28}	{0.47}	{0.47}
Control Mean	0.70	0.39	0.42	-0.13
Observations	149	149	149	149
Trust Script vs. No Script				
Trust Script	-0.18	0.01	-0.04	-0.20
	(0.08)	(0.08)	(0.08)	(0.19)
	[0.11]	[1.00]	[1.00]	[0.83]
	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.73	0.49	0.49	0.12
Observations	149	149	149	149

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Results in the *Income and Business Size (Objective and Complex)* Category

Table B.6: Complex Business Information

	Business Capital (Inv Hyp Sin)	Working Hours	Pay to HH Members (Inv Hyp Sin)	Hours Members No Pay	Pay to Employees (Inv Hyp Sin)	Revenue (Inv Hyp Sin)	Profits (Inv Hyp Sin)	Investment (Inv Hyp Sin)	Business Debts (Inv Hyp Sin)	Summary Index
Familiar vs. New Surveyor										
Familiar Surveyor	0.01	1.68	0.04	-0.66	-0.17	-0.08	-0.48	0.09	0.16	-0.04
	(0.07)	(1.71)	(0.07)	(1.22)	(0.10)	(0.07)	(0.20)	(0.13)	(0.17)	(0.06)
	[1.00]	[1.00]	[1.00]	[1.00]	[0.46]	[1.00]	[0.25]	[1.00]	[1.00]	[1.00]
	{1.00}	{0.78}	{0.78}	{0.78}	{0.63}	{0.74}	{0.31}	{0.78}	{0.78}	{0.78}
Control Mean	553.81	65.64	1.48	4.37	5.32	112.95	25.21	9.47	50.57	-0.05
Observations	827	835	834	834	835	824	821	811	836	836
In-Person vs. Phone Survey										
In-Person Survey	-0.03	-1.82	-0.16	-1.67	-0.05	-0.11	0.12	0.20	-0.02	-0.06
	(0.06)	(1.73)	(0.07)	(1.31)	(0.10)	(0.07)	(0.21)	(0.13)	(0.17)	(0.06)
	[0.96]	[0.64]	[0.24]	[0.64]	[0.96]	[0.55]	[0.96]	[0.55]	[0.97]	[0.64]
	{1.00}	{0.89}	{0.30}	{0.89}	{1.00}	{0.76}	{1.00}	{0.76}	{1.00}	{0.89}
Control Mean	518.74	66.51	1.79	4.48	4.76	105.10	19.18	9.51	55.06	-0.05
Observations	827	835	834	834	835	824	821	811	836	836
Trust Script vs. No Script										
Trust Script	-0.03	-1.76	0.01	1.53	-0.07	0.07	0.04	-0.17	-0.07	-0.03
	(0.07)	(1.71)	(0.07)	(1.20)	(0.10)	(0.07)	(0.21)	(0.13)	(0.17)	(0.06)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	559.37	66.99	1.58	3.03	5.02	109.19	21.02	10.91	57.64	-0.04
Observations	827	835	834	834	835	824	821	811	836	836

See Appendix Table B.2 for notes on estimation and hypothesis testing.

Table B.7: Total Household Income

	Household Income (Inv Hyp Sin)	Summary Index
Familiar vs. New Surveyor		
Familiar Surveyor	-0.29	-0.10
	(0.20)	(0.07)
	[0.17]	[0.17]
	{0.67}	{0.67}
Control Mean	45.00	0.03
Observations	836	836
In-Person vs. Phone Survey		
In-Person Survey	0.19	0.07
	(0.20)	(0.07)
	[0.55]	[0.55]
	{0.89}	{0.89}
Control Mean	41.37	-0.06
Observations	836	836
Trust Script vs. No Script		
Trust Script	-0.05	-0.02
	(0.20)	(0.07)
	[1.00]	[1.00]
	{1.00}	{1.00}
Control Mean	45.55	-0.02
Observations	836	836

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Results in the *Well-Being and Pro-Social Beliefs (Subjective and Non-Sensitive)* Category

Table B.8: Subjective Well-being

	Feel Happy	Feel Calm	Not Feel Sad	Summary Index
Familiar vs. New Surveyor				
Familiar Surveyor	-0.07	-0.05	-0.01	-0.11
	(0.03)	(0.03)	(0.03)	(0.06)
	[0.20]	[0.20]	[0.22]	[0.20]
	{0.15}	{0.20}	{0.37}	{0.19}
Control Mean	0.41	0.41	0.59	-0.01
Observations	834	836	835	836
In-Person vs. Phone Survey				
In-Person Survey	0.01	-0.00	-0.02	-0.01
	(0.03)	(0.03)	(0.03)	(0.06)
	[1.00]	[1.00]	[1.00]	[1.00]
	{0.76}	{0.77}	{0.54}	{0.77}
Control Mean	0.37	0.39	0.59	-0.05
Observations	834	836	835	836
Trust Script vs. No Script				
Trust Script	0.01	0.02	-0.03	0.00
	(0.03)	(0.03)	(0.03)	(0.06)
	[1.00]	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.36	0.37	0.60	-0.06
Observations	834	836	835	836

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Table B.9: Social Attitudes Toward General Topics

	People Can Be Trusted	People Are Fair In Business	People Are Helpful	Summary Index
Familiar vs. New Surveyor				
Familiar Surveyor	0.09	0.03	0.06	0.17
	(0.03)	(0.03)	(0.03)	(0.07)
	[0.02]	[0.15]	[0.07]	[0.03]
	{0.05}	{0.34}	{0.19}	{0.10}
Control Mean	0.31	0.52	0.59	-0.06
Observations	831	832	835	835
In-Person vs. Phone Survey				
In-Person Survey	-0.07	-0.03	-0.06	-0.16
	(0.03)	(0.04)	(0.03)	(0.07)
	[0.07]	[0.14]	[0.07]	[0.07]
	{0.14}	{0.34}	{0.19}	{0.14}
Control Mean	0.41	0.55	0.65	0.14
Observations	831	832	835	835
Trust Script vs. No Script				
Trust Script	-0.03	0.01	0.04	0.01
	(0.03)	(0.03)	(0.03)	(0.07)
	[1.00]	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.37	0.53	0.60	0.03
Observations	831	832	835	835

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Measures excluded from our pre-specified index construction

Table B.10: Knowledge of Refugees and Hosting Policy (for Ugandans)

	Allowed Outside Settlements	Know Intl Donations Shared w Ugandans	Summary Index
Familiar vs. New Surveyor			
Familiar Surveyor	0.02	0.05	0.10
	(0.04)	(0.04)	(0.07)
	[0.43]	[0.43]	[0.43]
	{0.37}	{0.25}	{0.25}
Control Mean	0.65	0.46	-0.11
Observations	687	687	687
In-Person vs. Phone Survey			
In-Person Survey	-0.06	-0.06	-0.17
	(0.04)	(0.04)	(0.08)
	[0.10]	[0.10]	[0.10]
	{0.19}	{0.21}	{0.14}
Control Mean	0.69	0.53	0.04
Observations	687	687	687
Trust Script vs. No Script			
Trust Script	-0.02	0.01	-0.01
	(0.04)	(0.04)	(0.07)
	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}
Control Mean	0.68	0.48	-0.04
Observations	687	687	687

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Results in the *Alignment With Prevailing Norms (Subjective and Sensitive)* Category

Table B.11: Support for Inclusive Refugee Hosting

	Supports Hosting	Freedom of Movement	Right to Work	Provide Land in Settlements	Provide Citizen-Ship	Accept More Refugees	Summary Index
Familiar vs. New Surveyor							
Familiar Surveyor	-0.02	-0.01	-0.00	-0.02	-0.03	-0.01	-0.12
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[0.93]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{0.36}
Control Mean	0.89	0.71	0.80	0.70	0.67	0.71	-0.01
Observations	687	687	687	687	836	836	836
In-Person vs. Phone Survey							
In-Person Survey	-0.02	-0.03	0.00	-0.01	-0.02	0.03	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.90	0.73	0.80	0.70	0.67	0.69	-0.04
Observations	687	687	687	687	836	836	836
Trust Script vs. No Script							
Trust Script	0.02	0.07	0.04	0.04	0.02	0.01	0.11
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)
	[0.69]	[0.31]	[0.53]	[0.53]	[0.94]	[0.94]	[0.40]
	{1.00}	{0.45}	{1.00}	{1.00}	{1.00}	{1.00}	{0.57}
Control Mean	0.87	0.67	0.79	0.68	0.64	0.69	-0.13
Observations	687	687	687	687	836	836	836

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Table B.12: Policy Preferences and Representation

	Satisfied w LC1	Satisfied w MP	Summary Index
Familiar vs. New Surveyor			
Familiar Surveyor	0.01	-0.03	-0.00
	(0.02)	(0.04)	(0.07)
	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}
Control Mean	0.85	0.63	0.04
Observations	812	656	814
In-Person vs. Phone Survey			
In-Person Survey	0.02	0.05	0.07
	(0.02)	(0.04)	(0.07)
	[0.88]	[0.88]	[0.88]
	{1.00}	{1.00}	{1.00}
Control Mean	0.84	0.60	-0.00
Observations	812	656	814
Trust Script vs. No Script			
Trust Script	0.00	0.02	0.02
	(0.02)	(0.04)	(0.07)
	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}
Control Mean	0.84	0.62	0.02
Observations	812	656	814

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Table B.13: Beliefs About Economic Effects of Immigrants (for Ugandans)

	Refugee Pos Effect Overall	Refugee Pos Effect Personally	Foreigner Pos Effect Overall	Foreigner Pos Effect Personally	Summary Index
Familiar vs. New Surveyor					
Familiar Surveyor	-0.06	-0.03	-0.03	0.04	-0.07
	(0.04)	(0.04)	(0.03)	(0.04)	(0.07)
	[0.62]	[0.62]	[0.62]	[0.62]	[0.62]
	{0.66}	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.61	0.57	0.76	0.58	-0.01
Observations	650	674	663	676	685
In-Person vs. Phone Survey					
In-Person Survey	-0.03	-0.04	0.01	-0.05	-0.10
	(0.04)	(0.04)	(0.03)	(0.04)	(0.08)
	[0.88]	[0.88]	[0.88]	[0.88]	[0.88]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.61	0.58	0.75	0.64	0.02
Observations	650	674	663	676	685
Trust Script vs. No Script					
Trust Script	0.01	-0.01	0.02	-0.00	-0.01
	(0.04)	(0.04)	(0.03)	(0.04)	(0.07)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.57	0.55	0.74	0.61	-0.05
Observations	650	674	663	676	685

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Table B.14: Social Attitudes toward Controversial Topics

	Refugee Pos Effect on Culture	Social Proximity Index (Ugandan)	Foreigner Pos Effect on Culture	Social Proximity Index (Refugee)	Cannot Beat Wife	Girl Can Refuse Marriage	People w HIV Can Work	Child under 15 Cannot Work	Child under 17 Cannot Work	Summary Index
Familiar vs. New Surveyor										
Familiar Surveyor	-0.05	0.01	-0.02	0.09	-0.06	-0.02	0.00	-0.01	-0.00	-0.15
	(0.03)	(0.07)	(0.04)	(0.16)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)	(0.07)
	[0.91]	[1.00]	[1.00]	[1.00]	[0.12]	[1.00]	[1.00]	[1.00]	[1.00]	[0.12]
	{0.97}	{1.00}	{1.00}	{1.00}	{0.09}	{1.00}	{1.00}	{1.00}	{1.00}	{0.12}
Control Mean	0.77	-0.01	0.31	-0.05	0.87	0.90	0.97	0.59	0.54	-0.01
Observations	643	687	652	149	831	835	834	836	829	836
In-Person vs. Phone Survey										
In-Person Survey	-0.01	-0.03	-0.06	-0.02	-0.05	0.00	0.01	-0.01	0.05	-0.04
	(0.03)	(0.07)	(0.04)	(0.15)	(0.03)	(0.02)	(0.01)	(0.03)	(0.04)	(0.07)
	[1.00]	[1.00]	[0.65]	[1.00]	[0.65]	[1.00]	[1.00]	[1.00]	[0.65]	[1.00]
	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}	{1.00}
Control Mean	0.75	0.03	0.32	0.03	0.86	0.87	0.96	0.58	0.54	-0.08
Observations	643	687	652	149	831	835	834	836	829	836
Trust Script vs. No Script										
Trust Script	-0.02	-0.04	0.07	-0.06	0.01	-0.04	-0.02	0.04	-0.01	-0.08
	(0.03)	(0.07)	(0.04)	(0.15)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)	(0.07)
	[1.00]	[1.00]	[0.23]	[1.00]	[1.00]	[0.23]	[0.23]	[0.58]	[1.00]	[0.59]
	{1.00}	{1.00}	{0.45}	{1.00}	{1.00}	{0.45}	{0.45}	{1.00}	{1.00}	{1.00}
Control Mean	0.75	0.02	0.26	0.04	0.84	0.90	0.98	0.56	0.56	-0.06
Observations	643	687	652	149	831	835	834	836	829	836

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.

Table B.15: Psychological Integration of Refugees (for Refugees)

	Feel Connected w Uganda	Not Feel Like Outsider	Summary Index
Familiar vs. New Surveyor			
Familiar Surveyor	-0.27	-0.21	-0.55
	(0.08)	(0.08)	(0.16)
	[0.00]	[0.00]	[0.00]
	{0.01}	{0.09}	{0.01}
Control Mean	0.64	0.51	-0.14
Observations	149	149	149
In-Person vs. Phone Survey			
In-Person Survey	-0.09	0.01	-0.04
	(0.08)	(0.08)	(0.16)
	[1.00]	[1.00]	[1.00]
	{1.00}	{1.00}	{1.00}
Control Mean	0.57	0.38	-0.40
Observations	149	149	149
Trust Script vs. No Script			
Trust Script	0.16	-0.09	0.04
	(0.08)	(0.08)	(0.16)
	[0.14]	[0.39]	[0.73]
	{0.45}	{1.00}	{1.00}
Control Mean	0.46	0.43	-0.43
Observations	149	149	149

Results estimated through ANCOVA regression with baseline controls selected through double-lasso. Heteroskedasticity-robust standard errors in parentheses. Brackets display sharpened q-values controlling the false discovery rate within the set of outcomes shown in this table (Anderson, 2008). Braces display sharpened q-values controlling the false discovery rate in the broader categories defined in Section 2.4.