Measuring the effects of survey methods on Women's Life Experience (WLE)

Thomas Assefe[1], Aditi Kadam[1], Nicholas Magnan[1], Ellen McCullough[1], Tamara McGovack [2] ([1]University of Georgia; [2]Grinnell College)

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August

Measuring the effects of survey methods on Women's Life Experience (WLE)

Thomas Assefa[1], Aditi Kadam[1], Nicholas Magnan[1], Ellen McCullough[1] and Tamara McGavock[3]

[1]University of Georgia

[2]Grinnell College

May 19, 2022

Abstract

Though there is a wide literature on women's empowerment, intimate partner violence (IPV), and gender norms, under-reporting and misreporting remain a major concern. Respondents may be reluctant to reveal honest answers because of social pressure, shame, taboo, fear of safety, or have normalized patriarchal norms and violence in their minds. Misreported answers can bias researchers' perceptions about the respondent's true beliefs about social norms. Systematic causally identified empirical evidence for the most effective ways of capturing sensitive behaviors and social norms is limited. In this paper, using a sample of 637 women from northern Amhara region in Ethiopia, we test whether survey methods that increase respondent privacy and anonymity affect the reporting of social norms (such as women's perceived independence, empowerment, and safety). We randomize the type of interview (phone vs in person), and gender of interviewer Additionally, we explore the effect of rapport conditional on interview type by comparing responses between women who have had prior frequent interactions with our study team and those who have not. Preliminary results show that when interviewed via mobile phone vs in-person, women report more of sensitive topics such as physical safety, emotional well-being, and freedom of movement. To our knowledge, this is the first paper that measures the effects of building a rapport with respondents on reporting social norms.

Keywords— gender, norms, beliefs, survey methods, international development

JEL Codes: C83, O12, J16

1 Introduction

Though there is a wide literature on women's empowerment, intimate partner violence (IPV), and gender norms, under-reporting and misreporting remain a major concern. When measuring sensitive topics that may embarrass or endanger the respondents if they report their true beliefs, it is likely that self-reported survey data may be misreported (Karlan & Zinman, 2008). As a mechanism to avoid embarrassment, participants may default to providing socially desirable answers. This misreporting can bias researchers' perceptions about the respondent's true beliefs of social norms. These biases may vary systematically with certain interview types (e.g. in person in front of her husband), interviewer identities (e.g. men), or preexisting relationships (e.g. unknown stranger).

In this paper, we measure the effect of interview type and of interviewer on measures of women's perceived independence, empowerment, and safety. Literature on best practices in interviewing respondents on sensitive issues recommends obtaining complete privacy of the respondent. Given that most survey enumerators are strangers, it is difficult to execute control over the presence of other household members. Obtaining complete privacy in face-to-face (FTF) interviews can also be difficult in households with limited space where there is a chance that others may be able to hear the interview. Other interviewer characteristics such as gender, tone of communication, non-verbal gestures, can also affect responses. These interviewer characteristics can generate non-random bias in responses (e.g. respondents may feel more comfortable answering to enumerators of the same gender). We note that for questions that may appear threatening or anxiety inducing (for e.g. asking women about their experiences with emotional violence), the bias only goes in one-direction: under-reporting. Thus, in the absence of documented truth, we assume that higher reporting of contra-normative behavior is indicative of lower socially-desirable response bias (Bradburn et al.,

1978). We acknowledge that this may not be true for all respondents, but is generally the case assumed in the literature on reporting of sensitive behaviors (Bradburn et al., 1978; Schroder et al., 2003; Locander et al., 1976; Krumpal, 2013).

Given the existence of reporting biases, there is a need to causally identify the effects of different survey instruments to measure sensitive beliefs such as social norms and behaviors of violence, especially in a rural, developing country context. Using a sample of 637 women from rural Amhara region in Ethiopia, we test whether survey methods that increase respondent privacy and anonymity affect the reporting of social norms compared to the status quo in person, face-to-face interviews. We randomly assign each female to answer the same questionnaire in one of three different ways: (1) via in person interview with a field enumerator (all field enumerators are male); (2) via phone interview with a male enumerator; (3) via phone interview with a female enumerator. Some of our respondents are also randomly assigned to frequently interact with the enumerators over the phone (as part of a different treatment of the "Her Time" project which is independent of WLE treatments). This survey design allows us to examine the effect of interview method conditional on a built-up rapport, and the effect of rapport conditional on interview type.

Preliminary results indicate that interviewing over mobile phone results in higher reporting of conservative social norms. We interpret this as less under-reporting of true social beliefs. Conditional on using a mobile phone, we find no significant effect of gender of the interviewer on the respondent's reported beliefs about physical safety or emotional well-being. Additionally, we find that building a rapport may reduce social desirability bias for questions associated with strong patriarchal beliefs.

Our paper contributes to the growing literature on alternative survey methods to study sensitive behaviors inlcuding IPV (Blattman et al., 2016; Karlan & Zinman, 2012; Blair et al., 2018; Chuang

et al., 2021; Bulte & Lensink, 2019). It is known that systematic measurement error is a large and hard to address problem, which makes it imperative to test new methods of data collection (Blattman et al., 2016). The act of asking sensitive questions such as beliefs about IPV can place a high emotional burden on both the enumerators and the respondents. Alternative methods such as Audio Computer Assisted Self-Interviewing (ACASI) have been used as self-interviewing tools instead of the usual in-person interviews (Cullen, 2020; Fincher et al., 2015; Park et al., 2021). There have been mixed results about whether self-interviewing tools are useful in reducing under-reporting. Park et al. (2021) find that self-interviewing spuriously increases reported IPV due to the complicated nature of ACASI that leads to misunderstanding of the questions. Since our methodology involves human interaction which can reduce incorrect interpretation of questions, we add to this literature by over-coming the caveats of ACASI, while still giving autonomy to respondents to feel comfortable answering sensitive questions. To our knowledge, this is the first paper that measures the effects of building a rapport with respondents on reporting social norms.

This paper also contributes to the recent literature on survey methodologies that are aimed at preserving respondent confidentiality [1]. Due to the emotionally heavy nature of sensitive questions such as physical violence, there is a need to find methods that maximize the respondents willingness to provide truthful responses. In our paper, we implement three ways of making respondents feel comfortable to provide true answers- phone, gender, and rapport. Estimating the effects of using mobile phone surveys for sensitive behavior are especially relevant now, given the changing methods of survey implementation remotely since world-wide lock downs due to COVID-19. If we find that mobile phone surveys reduce misreporting, then this finding will prove to be a very important cost reducing strategy for survey data for both sensitive and non-sensitive behaviors.

[1] see Höglinger & Jann (2018) for a review of randomized response techniques (RRT)

2 Data and Experimental design

This paper is a part of a project lead by McCullough & McGavock (2020) titled "Her Time: A Time Use Study of Women Participating in Livelihoods Programs in Ethiopia." The survey design of this experiment leverages the high-frequency data collection of the "Her Time" project. The sample consists of 900 eligible women from poor households in the northern Amhara region in Ethiopia [2]. Using program enrollment records, we draw a random sample of Village Economic and Social Associations (VESA) groups in the Meket and Wadla *woredas* [3] of the North Wollo zone. Within each VESA selected according to the selection criteria described above, we recruit one woman from each unique household in the VESA, regardless of whether the woman is a VESA participant herself or not. The recruited female respondents are randomly assigned to one of three possible treatment arms, which determine the method to be used to interview the woman (both types of data collection, in-person and phone, occur during the 7-day period following the conclusion of baseline interviews within each VESA). During baseline interviews, household characteristics including livelihoods activities were collected. During the baseline study, we provided information to inform women's consent in the study. The women assigned the phone treatment group received their cell phones and instructions on their use.

The WLE survey module incorporates some questions used by the Demographic and Health Surveys (DHS) that aim to understand women's safety and sense of empowerment in their relationships. Table A1 explains the questions asked. options for each question, and how the question is fed

[2]The sample is randomly selected from the population of eligible households who choose to participate in a safety net program rather than the larger population of eligible households in the region. In order to be eligible for the programs, members must qualify for and participate in Ethiopia's Productive Safety Net Program (PSNP), a government safety net program that targets poor and vulnerable households in chronically food insecure kebeles (districts) in Ethiopia.

[3] *Woredas* are tertiary administrative units similar to districts, while kebeles comprise the next level of administrative units, similar to blocks or sub-districts.

into the index (negatively or positively). We include questions about women's perceptions generally speaking about whether men are justified in hitting their wives under different circumstances, women's perceived ability to insist on using condoms with a partner known to have an STI. For questions on Physical safety, we ask the women's beliefs about spousal violence. We ask whether the women thinks that it is justified for *a husband* to beat his wife when she goes out without telling him, neglects the children, argues with him, refuses to have sex with him, or burns the food. These are similar to the DHS module on beliefs about physical violence. For emotional well-being and freedom of movement questions, we ask the women about her *own experiences* with her spouse/household members. We ask her whether respect in her relationship is mutual, she believes her husband acts in her best interest, she feels comfortable being vocal about a disagreement, his health is more important than her own health, and if she is afraid of her current (or last) partner. We also ask if she feels that her travelling outside the household is restricted for various purposes such as markets, family, hospitals, community meetings, religious spaces, and spaces outside of her community/village.

We have three treatment groups (six treatment bins) for the WLE survey (in-person by male enumerator, over the phone by male enumerator, and over the phone by female enumerator). We did not conduct in-person interviews with female enumerators given security concerns about remote travel for women in Ethiopia. This is often the case for conducting household surveys in rural areas. The repeated contact by phone for the phone group (7-10 days, four to five times per day plus two in-person interviews) prior to the WLE module may cause women to feel some level of rapport with the survey team (and perhaps specifically with the female call center enumerators) than those assigned to the control group (with only two in-person interviews) do not. The following table depicts how the full sample is split across the six WLE treatment bins.

Table 1: Treatment groups

	In-person, Male	Phone, male enumerator	Phone, female enumerator	Total
No-Rapport	135	60	90	285
Rapport	241	142	120	503
Total	376	202	210	788

3 Empirical Strategy

To measure the effect of different survey methods, we estimate the equation-

$$y_{ik} = \beta_1 + \beta_2 Phone_{male_i} + \beta_3 Phone_{female_i} + \beta_4 PX_i + X_i'\beta_5 + \epsilon_{ik} \quad (1)$$

where y_{ik} refers to outcome k that is part of the WLE domain and X is a vector of control variables. $Phone_{male}$ is a dummy for phone interview conducted by a male enumerator and $Phone_{female}$ is a dummy for phone interview conducted by a female enumerator. The omitted dummy variable is for in-person interview conducted by a male. Therefore, β_2 is the impact of moving from an in-person interview conducted by a male to a phone survey conducted by a male. The null hypothesis $\beta_2 = 0$ tests that there is no difference in the outcome whether a survey is conducted by a male in-person or by a male over the phone. β_3 is the impact from moving from an in-person survey conducted by a male to a phone survey conducted by a female. The null hypothesis $\beta_2 = \beta_3$ is that for a phone survey, it does not matter if an interview is conducted by a male or a female. Variable PX represents the impact of having built a rapport with the respondent over the phone during previous calls as part of the other unrelated treatments across all WLE treatment groups (Male in-person, male by phone, and female by phone). We control for this since respondents could have different answers if there is a pre-developed rapport with the enumerators.

We will test the following three hypothesis-

- Hypothesis 1: For a male enumerator, there is no difference in WLE responses between an in-person or a phone survey. ($\beta_2 = 0$)

- Hypothesis 2: For a phone survey, there is no difference in WLE responses for a survey conducted by a male versus female enumerator. ($\beta_2 = \beta_3$)

- Hypothesis 3: As an extension to the previous model, we can explore whether building a rapport between the survey team and respondents (i.e. by contacting the respondents many times via phone surveys prior to asking the WLE questions) affects the responses to WLE questions. In other words, will a female enumerator get different answers to such sensitive questions when a respondent has already been communicating with the survey team regularly about her daily life? To do this, we can estimate the following specification (and the analogs for the other WLE survey types):

$$y_{ik} = \beta_1 + \beta_2 PX_i + X'\beta_3 | Phone_{female_i} = 1 \tag{2}$$

  The null hypothesis ($\beta_2 = 0$) suggests that there is no difference in WLE question responses for respondents who have had more frequent contact with the survey team.

- We also test an additional hypothesis for the rapport effect by testing whether developing a rapport with the survey team as a whole has an effect on responses to WLE questions. We do this by removing the restriction on the sample that were only surveyed by female enumerators. We test the following specification-

$$y_{ik} = \beta_1 + \beta_2 PX_i + X'\beta_3 \tag{3}$$

As in any survey about sensitive questions, non-response rates will be considered an outcome with potential to be impacted by treatment assignment. Beyond response rates, we will also compare women's responses to the questions across these six treatment groups, by constructing an index based on women's responses to the various questions in these modules. The primary outcomes of interest can be described at both the extensive and the intensive margins. At the extensive margin, we can test the response rate, i.e. how many women choose to answer the questions on beliefs on gender norms when they are interviewed by phone or in-person and by a male enumerator. Conditional on answering, we can then test the difference in the reporting for the sensitive behavioral norm, e.g., women reporting "yes" for justification of husband beating behavior or "most of the time" for permission to travel alone to local places like the market. Following Anderson (2008), we will create an outcome variable for the three separate WLE domains listed below that is an average of the outcomes measured within that domain weighted by the inverse covariance matrix of the outcomes within that domain. Table A1 in the Appendix shows the list of questions that were asked within each outcome group, and how the responses were coded for index creation.

1. Norms related to physical violence: We use the outcomes from two questions related to norms about physical violence. The women answer from a binary set of outcomes "yes" or "no" for each of these questions.

   (a) Whether the woman thinks it is justified that a husband beat his wife when:

   - she goes out without telling him (yes/no)

- she neglects the children (yes/no)
- she argues with him (yes/no)
- she refuses to have sex with him (yes/no)
- she burns the food (yes/no)

(b) Whether the woman is of the opinion that it is the woman's role to tolerate violence in her household in order to keep her family together (yes/no)

2. Norms related to emotional and physical well-being: These outcomes are related to how the woman feels about her spouse and how she thinks her spouse feels about her. Where responses follow a multi-point Likert scale, we will create multiple outcome variables based on whether the answer exceeds the ordered response, from the 2nd lowest to the highest.

   (a) Whether the woman respects her husband (4-pt)

   (b) Whether she feels that her husband respects her (4-pt)

   (c) Whether she trusts her husband to do things that are in her best interest (4-pt)

   (d) When she disagrees with her husband, whether she feels comfortable telling him that she disagrees (4-pt)

   (e) Whether it is her opinion that her husbands health is more important than her health (yes/no)

   (f) Whether she is afraid of her current (or last) husband/partner (3-pt)

3. Norms related to freedom of movement: These outcomes are related to the woman's freedom to go to places by herself. For the following outcomes, women select between three options:

most of the times, sometimes, or rarely. Where responses follow a multi-point Likert scale, we will follow the procedure described above.

(a) To what extent the woman's husband/partner or another household member objects her going alone to (3-pt answer):

- an urban center
- the market
- visit family or relatives
- visit a friend / neighbor's house
- the hospital / clinic / doctor (seek health service)
- temple / church / mosque
- a public village gathering / community meeting
- training for NGO or programs
- outside her community or village

(b) Whether her spouse knows where she goes/is (3-pt)

For all these questions, women have to pick between three options: most of the times, sometimes, or rarely.

We create indices for physical violence, emotional violence and freedom of movement. We follow Anderson (2008) and Haushofer et al. (2019) who use this approach to develop an IPV index, to create standardized inverse covariance weights of sub-indicators for each of the three domains described above. First, we recode the outcomes where a high value corresponds with more empowerment, so that high values universally correspond with lower levels of empowerment

/ higher levels of IPV. Next, we compute a covariance matrix for each outcome within the outcome group, then we invert the matrix and create weights based on the row sum corresponding to each outcome. Finally, we normalize each outcome variable by demeaning it and dividing it by the control group standard deviation, and weight it. More information about index creation can be found in the Appendix.

4 Results

Results for equations (1), (2), and (3) can be found in Table 1. Columns (1) and (2) show results from equation (1). Columns (3) and (4), show results from equations (2), and (3), respectively. We find that using a phone significantly reduces the likelihood of over-reporting empowerment. In other words, when asked on the phone, women are less likely to state responses that indicate higher empowerment. The result is significant for all three outcome groups. This effect is strongest for questions about physical safety. Respondents are 0.34 standard deviations less likely to have beliefs that indicate their own physical well-being i.e. they are more likely to say that they think it is justified for man to commit acts of violence by when asked on the phone, compared to when asked in-person by a male enumerator. Similarly, when asked on the phone, respondents are 0.2 standard deviations less likely to state that feel emotionally secure in their relationships, and 0.24 standard deviations less likely to state that have freedom to go where they please (without any objections from other family members).

Conditional on using a phone, we find little to no significant effect of the gender of the enumerator. The only significant results we find are for freedom of movement. We find that when asked by male enumerators, respondents tend to be more likely to state that they have freedom to

move around without the approval of other family members. Lastly, building rapport with female enumerators seems to increase responses that indicate a higher emotional and physical well-being by 0.128 standard deviations. Having built a rapport also increases the likelihood of stating beliefs that indicate better physical safety by 0.266 standard deviations. These results for rapport effect are surprising. We expected, similar to phone effects, that building a rapport would reduce under-reporting by decreasing the reporting of higher empowerment. In contrast, we find that rapport increases reporting of empowerment. One explanation for this could be that without having built a rapport, respondents feel safer responding with socially desirable responses. In areas where social norms dictate lower levels empowerment, women might resort to self-reporting lower levels of empowerment because of the fear of facing "backlash" from the society and/or family members. For example, if the norms impose that it is "normal" for men to physically abuse women, then the respondents would give responses that indicate higher compliance with abuse. However, after establishing a rapport, the respondents feel confident confiding in their true response without having the fear of deviating from the social norms.

Table 2: WLE effects

	(1) Phone Effect	(2) Gender Effect	(3) Rapport Effect (Female enumerators only)	(4) Rapport Effect (Full sample)
Physical Safety Index	-0.373***	-0.153	-0.099	0.2442***
	(0.0860)	(0.074)	(0.080)	(0.069)
Emotional Well-being Index	-0.208***	0.108	0.0772	-0.0164
	(0.076)	(0.056)	(0.0516)	(0.0635)
Freedom of movement Index	-0.229***	0.0279	-0.0669	-0.0177
	(0.082)	(0.074)	(0.0527)	(0.0645)
Observations	637	637	220	637

standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5 Discussion

Studying sensitive topics is complicated due to the controversial nature of the topics and the taboos and stigmas attached to revealing honest preferences. In this paper, we target the collection better data on beliefs about social norms, emotional well-being of respondents, and levels of empowerment as indicated by freedom of making independent decisions. We randomize the method of interview (phone vs. in person), gender of the enumerator, and building a rapport to estimate effects of using alternative survey mythologies on reporting of sensitive beliefs. The most important insight from our findings is that using mobile phones for asking questions is an effective in reducing the under-reporting of beliefs about social norms and empowerment. Using mobile phones reduces the emotional burden of having to answer sensitive questions face-to-face to a stranger, gives respondents higher privacy, and increases safety of respondents by reducing chances of being over-heard by another family/society member. Estimating the effects of using mobile phone surveys for sensitive behavior are especially relevant now, given the changing methods of survey implementation remotely since world-wide lock downs due to COVID-19.

We also find preliminary evidence that building a rapport can help in reducing socially desirable answers in areas with strongly ingrained ideas of male dominance and patriarchy. We find effective alternative survey measurement methods that are less biased, logistically easier and cheaper, and more ethical than the status-quo. These methods reduce the overall costs (monetary and non-monetary) of asking sensitive questions. A reduction of costs increases the likelihood of adoption by more researchers and contribute to new research on sensitive topics such as IPV and social norms, especially in the developing context. Better data means an improvisation in our understanding of the causes of IPV and may broaden the range of interventions that could be used to reduce it

(Peterman et al., 2018).

References

Blair, G., Coppock, A., & Moor, M. (2018). When to worry about sensitivity bias: evidence from 30 years of list experiments. *Work. Pap., Univ. Calif., Los Angeles, CA Google Scholar Article Location*.

Blattman, C., Jamison, J., Koroknay-Palicz, T., Rodrigues, K., & Sheridan, M. (2016). Measuring the measurement error: A method to qualitatively validate survey data. *Journal of Development Economics*, *120*, 99–112.

Bradburn, N. M., Sudman, S., Blair, E., & Stocking, C. (1978). Question threat and response bias. *Public opinion quarterly*, *42*(2), 221–234.

Bulte, E., & Lensink, R. (2019). Women's empowerment and domestic abuse: Experimental evidence from vietnam. *European economic review*, *115*, 172–191.

Chuang, E., Dupas, P., Huillery, E., & Seban, J. (2021). Sex, lies, and measurement: Consistency tests for indirect response survey methods. *Journal of Development Economics*, *148*, 102582.

Cullen, C. (2020). Method matters: Underreporting of intimate partner violence in nigeria and rwanda. *World Bank Policy Research Working Paper*, (9274).

Fincher, D., VanderEnde, K., Colbert, K., Houry, D., Smith, L. S., & Yount, K. M. (2015). Effect of face-to-face interview versus computer-assisted self-interview on disclosure of intimate partner violence among african american women in wic clinics. *Journal of interpersonal violence*, *30*(5), 818–838.

Höglinger, M., & Jann, B. (2018). More is not always better: An experimental individual-level

validation of the randomized response technique and the crosswise model. *PloS one*, *13*(8), e0201770.

Karlan, D., & Zinman, J. (2008). Lying about borrowing. *Journal of the European Economic Association*, *6*(2-3), 510–521.

Karlan, D. S., & Zinman, J. (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics*, *98*(1), 71–75.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & quantity*, *47*(4), 2025–2047.

Locander, W., Sudman, S., & Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association*, *71*(354), 269–275.

McCullough, E., & McGavock, T. (2020). Her time: a time use study of women participating in livelihoods programs in Ethiopia. *AEA RCT Registry. June 04. https://doi.org/10.1257/rct.5961-1.0.*.

Park, D. S., Aggarwal, S., Jeong, D., Kumar, N., Robinson, J., & Spearot, A. (2021). Private but misunderstood? evidence on measuring intimate partner violence via self-interviewing in rural liberia and malawi. Tech. rep., National Bureau of Economic Research.

Peterman, A., Palermo, T. M., Handa, S., Seidenfeld, D., & Team, Z. C. G. P. E. (2018). List randomization for soliciting experience of intimate partner violence: application to the evaluation of zambia's unconditional child grant program. *Health economics*, *27*(3), 622–628.

Schroder, K. E., Carey, M. P., & Vanable, P. A. (2003). Methodological challenges in research on sexual risk behavior: Ii. accuracy of self-reports. *Annals of behavioral medicine*, *26*(2), 104–123.

6 Appendix

Table A1: Index creation

Question	Answer options	Measurement	Coding to index
Physical Safety Index			
1. Whether the woman thinks it is justified that a husband beat his wife when:	yes/no	=1 if yes, = 0 if no	negatively coded
a. she goes out without telling him			
b. she neglects the children			
c. she argues with him			
d. she refuses to have sex with him			
e. she burns the food			
2. Whether the woman is of the opinion that it is the woman's role to tolerate violence in her household in order to keep her family together	yes/no	=1 if yes, = 0 if no	negatively coded
3. Whether the woman is of the opinion that a woman is justified in refusing sex with her husband if she knows he has sex with other women?	yes/no	=1 if yes, = 0 if no	positively coded
Emotional and Physical Well-being Index			
1. Whether the woman respects her husband		4-pt Likert scale	positively coded
2. Whether she feels that her husband respects her		4-pt Likert scale	positively coded
3. Whether she trusts her husband to do things that are in her best interest		4-pt Likert scale	positively coded
4. When she disagrees with her husband, whether she feels comfortable telling him that she disagrees		4-pt Likert scale	positively coded
5. Whether it is her opinion that her husbands health is more important than her health	yes/no	=1 if yes, = 0 if no	negatively coded
6. Whether she is afraid of her current (or last) husband/partner		3-pt Likert scale	negatively coded
Freedom of Movement Index			
1. To what extent the woman's husband/partner or another household member objects her going alone to:		3-pt Likert scale	negatively coded
a. an urban center			
b. the market			
c. visit family or relatives			
d. visit a friend / neighbor's house			
e. the hospital / clinic / doctor (seek health service)			
f. temple / church / mosque			
g. a public village gathering / community meeting			
h. training for NGO or programs			
i. outside her community or village			
2. Whether her spouse knows where she goes/is		3-pt Likert scale	negatively coded

6.1 Index creation

We create indices for physical safety, emotional and physiological safety, and freedom of movement as follows. We follow Anderson (2008) to create standardized inverse covariance weights (ICW) of subindicators for physical safety, emotional and physiological safety, and freedom of movement.

We will do that as follows:

First, for each outcome variable y_{jk} where j indexes the outcome group and k indexes variables within that outcome group, we will re-code the variables such that high values correspond to higher empowerment/lower experiences of violence. Then, we create the index using the following steps-

1. First, we compute the covariance matrix $\hat{\Sigma}_j$ for outcomes in outcome group j, which will be calculated as:

$$\hat{\Sigma}_{jmn} = \sum_{i=1}^{N_{jmn}} \frac{y_{ijm} - \bar{y}_{jm}}{\sigma^y_{jm}} \cdot \frac{y_{ijn} - \bar{y}_{jn}}{\sigma^y_{jn}} \tag{4}$$

   Where N_{jmn} is the number of non-missing observations for outcomes m and n in outcome group j, $\bar{y}_{jm}$ and $\bar{y}_{jn}$ are the means for outcomes m and n, respectively, in outcome group j, and σ^y_{jm} and σ^y_{jn} are the standard deviations in the pure control group (which in our case is the group assigned to the "in-person, male enumerator") for the same outcomes.

2. Then, we invert the covariance matrix and define weight w_{jk} for each outcome k in outcome group j by summing entries in the row of the inverted covariance matrix corresponding to that outcome.

$$w_{jk} = \sum_{l=1}^{K_j} c_{jkl} \tag{5}$$

   Where K_j is the total number of outcome variables in outcome group j.

3. Lastly, we transform each outcome variable by subtracting its mean and dividing by the pure control group standard deviation, and then weight it with the weights obtained from above.

The result $\hat{y}_{ij}$ is the generalized least squares estimator (Anderson, 2008).

$$\hat{y}_{ij} = \left(\sum_{k \in K_{ij}} W_{jk} \right)^{-1} \sum_{k \in K_{ij}} w_{jk} \frac{y_{ijk} - \bar{y}_{jk}}{\sigma^{y}_{jk}} \tag{6}$$

Where K_{ij} denotes the set of non-missing outcomes for observation i in outcome group j, and $W_{jk} = \sum_{k \in K_{ij}} w_{jk}$

Table A2: Results from equation 1

	Physical Safety Index			Emotional Well-being Index			Freedom of Movement Index		
	Phone Effect	Rapport Effect	Rapport Effect 2	Phone Effect	Rapport Effect	Rapport Effect 2	Phone Effect	Rapport Effect	Rapport Effect 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Phone Male	-0.373***			-0.235***			-0.229***		
	(0.0860)			(0.0808)			(0.0821)		
Phone Female	-0.220**			-0.337***	-0.257***				
	(0.0868)			(0.0631)			(0.0666)		
Rapport	0.255***	-0.0189	0.0772	-0.0189	0.0772	-0.0164	-0.0161	-0.0669	-0.0177
	(0.0694)	(0.118)	(0.0693)	(0.0635)	(0.0516)	(0.0635)	(0.0643)	(0.0527)	(0.0645)
Constant	0.149	0.0544	-0.0609	-1.072***	-1.475***	-1.245***	1.911***	0.288	1.760***
	(0.266)	(0.473)	(0.259)	(0.256)	(0.283)	(0.241)	(0.285)	(0.183)	(0.284)
Observations	637	220	637	637	220	637	578	190	578
R-squared	0.076	0.042	0.048	0.197	0.610	0.165	0.514	0.255	0.502

standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$