Self-Administered Field Surveys on Sensitive Topics

April 9, 2020

Online Appendix

A   Survey Procedures

In each of the study's barangays, we randomly selected names from the list of registered voters.[1] Enumerators went to each barangay to attempt to interview the selected subjects. If they were able to locate the subject's household but could not interview the subject (for example if the subject works in Manila, or if she declined to participate), the enumerator interviewed another adult living in the selected household instead. If the household could not be located, the enumerator moved on to the next name on the list of randomly-selected individuals. They repeated this process until they had interviewed 15 subjects in each barangay. In total, 32.8% of subjects were "proxy" individuals living in the selected subject's household. Of these, the overwhelming majority were the spouse or other close relative of the selected respondent.

We selected the sensitive question about reporting insurgent activity using the following procedures:

1. We first created a list of eight potential questions which we believed (a) measured substantively important topics and (b) fit the criteria of addressing a "taboo" topic without placing subjects' or enumerators' safety at risk.

2. We showed this list to two other experts on Filipino politics who have lived and worked in the Philippines and explained the criteria for selection. After some discussion with these experts, we narrowed the list to three possible questions:

   a. Do you know someone who is a member of an anti-government group?

   b. If you knew about the activities of an anti-government group, would you report them to the authorities?

   c. In the Philippines, it is illegal to sell your vote. Did you receive money in return for your vote in the most recent elections?

3. We piloted these three questions to 60 subjects using a convenience sample of Filipinos known to our enumerators.

---

[1]We estimate that about 85% of eligible adults are registered to vote.

4. Our most experienced enumerators recruited subjects with whom they had a prior relationship, including family members, neighbors, and friends, explained that they wish to pilot a survey question, and asked them the three sensitive questions using one of the three response methods. Enumerators then debriefed the subjects by explaining the goals of the study and asking for their reactions to the question, including whether they thought it was too sensitive and should not be asked of the general population, whether they were confused by the method of questioning, and whether.

5. Based on the feedback from the pilot survey, we selected the question about reporting insurgent presence to the authorities.

Figure 1: Pilot Survey Summary Statistics

	Know member of an anti-gov't group	Would report anti-gov't activities to authorities	Sold vote	None
It is understandable if any of these questions made you feel uncomfortable. Did any of them make you so uncomfortable that you think we should not ask them of people?	21.67%	11.67%	5.00%	61.67%
Do you believe that any of these questions are dangerous to ask? Remember that responses are completely anonymous and will only be shared with the researchers.	26.67%	11.67%	3.33%	58.33%

	It made me confused	It was not confusing
How did you feel about using the tablet to **enter your answers?**	2.86%	97.14%
How did you feel about answering the question based on the **coin toss?**	18.75%	81.25%

	Feel Safer	Feel More Worried	No Difference
How did you feel about using the tablet to **enter your answers?**	76.32%	5.26%	18.42%
How did you feel about answering the question based on the **coin toss?**	45.24%	11.90%	42.86%

Ethical Considerations

While the question about reporting insurgents is sensitive, it posed minimal risk to subjects' physical safety. After extensive consultation with local researchers, the Philippine National Police, our institution's IRB, and other local stakeholders, we determined that the likelihood of identifiable responses falling into the hands of the NPA or the police was so low, and the likelihood of either group retaliating against subjects in such an event was also so low, that the risks were far outweighed by the potential benefits of the research. Data remained secure throughout the study, and we are not aware of any adverse experiences by our subjects or enumerators due to the survey. Furthermore, we were conscious of the fact that a subset of respondents would be asked to respond directly and verbally. Thus, while we sought a question on a sensitive topic, we only considered questions that would not make insurgents or police so angry as to risk a violent reaction of they observed the response *in our specific setting*. While there are likely places in the Philippines in which insurgent-government violence is so severe that our question might have elicited a violent response, conflict in Sorsogon Province is relatively low-grade, and we conducted the research in barangays with relatively little insurgent activity.

We conducted both the pilot and the main survey with approval from our institution's IRB. Study participants provided informed consent, which included the following language:

> "If you agree to participate, we will ask you some questions about the government, anti-government groups, and crime, including your personal attitudes towards these groups and issues."

As well as,

> "I'd like to ask you about your attitudes towards sensitive topics, including the government's fight against anti-government groups and your experience with illegal activities. These questions sometimes make people uncomfortable. You always have the option of refusing to answer a question, and you may stop the interview at any time. If you admit to illegal behavior, it is possible that you could face legal consequences if your answers become known to the authorities. While we do not intend to share your answers with anyone, you should be aware of this risk. Because this is a research study, there may also be some unknown risks that are currently unforeseeable."

After the survey concluded, we found that just under 1/4 of interviews occurred in the presence of a bystander, according to enumerator coding. While we anticipated some bystander presence, we did not expect the degree to which this would occur. In our enumerator protocol manual and during in-person enumerator training, we specified that if bystanders were in the area, enumerators should ask respondents to move to a more private area, such as outside the house. We also instructed enumerators to end the survey immediately if they felt that the respondent or others near the interview area were acting hostile in any way. Based on PI debriefings with enumerators, nearly all surveys began in a private location. However, in some cases, bystanders (almost always members of the household) would return to "hover" at a distance or check in with the respondent about daily

tasks. In these cases, enumerators would reiterate the importance of conducting the survey in private. We instructed enumerators to err on the side of indicating that bystanders were present if they felt that at any point during the survey the respondent could have expected a bystander to overhear their answers.

A clear lesson learned from this experience is the need for even more thorough training of enumerators about the importance of privacy. The presence of bystanders is potentially harmful not just for the data quality issues we discuss below, but also to ensure the safety and comfort of both respondents and enumerators.

In our study, the main concern from a breach of privacy was that sensitive responses could fall into the hands of the authorities or insurgents, and that these groups might retaliate against respondents who gave the "wrong" answer. As we discuss above, we did not believe retaliation was likely even if privacy were breached. Two additional factors mitigate the ethical issues posed by bystanders.

First, respondents were aware of bystanders' presence when they decided whether to participate in the survey or respond to a question, and they had the opportunity to decline to participate or respond. While social pressure may have made some respondents uncomfortable, our finding that participants were significantly less likely to answer the sensitive question when bystanders were present indicates a substantial amount of agency by respondents (while reinforcing the notion that it would have been better not to place them in that position in the first place). In this sense, respondents' awareness of bystanders differentiates this threat from the issues of post-collection data security, which would occur after respondents have lost all agency over their participation.

Second, we believe that the presence of bystanders usually constituted a small number of family members or friends already in the respondent's house when the interview began, based mainly on debriefing enumerators but also on a small number of PI observations during piloting. We presume that family and friends would be unlikely to reveal respondents' answer choices by reporting information to the PNP or NPA if they believed it would harm the respondent.

To reiterate, we did not expect that such a large number of interviews would occur with bystanders present. While the above characteristics mitigate ethical concerns, these concerns could be eliminated entirely by ensuring that no one except the respondent and enumerator were present during the interview. Researchers should strive to conduct interviews in private. While no harm came to any of our respondents or enumerators as a result of participating in our study, the prevalence of bystanders demonstrates a need for more thorough training of enumerators on this issue, and a specific procedure for how enumerators should behave if a private location cannot be found.

B   Effects on Response Rates and Answer Choice

Response Rates: Regression models in Table 1 show that for the sensitive question, the difference across methods reported in Figure 2 are statistically significant. The dependent variable is a dichotomous indicator coded "1" if the respondent declined to answer, otherwise "0." We use logistic regression with standard errors clustered by barangay. Models 2 and 3 include barangay fixed effects. Model 3 shows that differences persist after controlling

for whether onlookers were present during the interview (*Bystander*), and the respondent's gender, age, education, and income.[2]

Table 1: Non-Response to Sensitive Question

	(1)	(2)	(3)
Self	-0.169**	-0.179**	-0.181*
	(0.0860)	(0.0860)	(0.0991)
Forced Choice	-0.435***	-0.454***	-0.474***
	(0.0829)	(0.0891)	(0.0900)
Bystander			0.302***
			(0.109)
Self x Bystander			-0.0605
			(0.178)
Male			-0.0466
			(0.0805)
Age			0.00190
			(0.00262)
Education			0.114***
			(0.0319)
Income			-0.00000116
			(0.00000419)
Constant	-0.895***		
	(0.0663)		
Barangay FE	No	Yes	Yes
Observations	4502	4217	4141

Logistic regression with barangay-clustered SE.

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 2 tests the effects of self enumeration (relative to direct questioning) on answer choice using logistic regression. While the relationship between self enumeration and affirmative responses is negative, it is not statistically significant. In the next section, we show that self enumeration also has no significant effect on the question about high school graduation.

Model 3 again includes an interaction with Bystander to test whether self-enumeration

---

[2]Despite enumerators' best efforts, 24.2% of surveys experienced bystanders.

Table 2: Would report anti-government group to the police

	(1)	(2)	(3)
Self	-0.0142	-0.0588	-0.115
	(0.0949)	(0.0934)	(0.111)
Bystander			-0.284*
			(0.171)
Self x Bystander			0.176
			(0.233)
Male			0.175
			(0.107)
Age			-0.0169***
			(0.00354)
Education			0.138***
			(0.0443)
Income			0.00000199
			(0.00000634)
Constant	0.456***		
	(0.0680)		
Barangay FE	No	Yes	Yes
Observations	2217	2075	2016

Logistic regression with barangay-clustered SE.

* $p < .10$, ** $p < .05$, *** $p < .01$

affects responses only when there are onlookers. The interaction term is not significant. However, the negative relationship between *Bystander* and willingness to report insurgents hints at a potential problem with our design. We expected that the enumerator's presence should *inflate* affirmative responses, as respondents likely viewed our enumerators as authority figures. Our enumerators were, on average, better educated than the average Sorsogon resident, wore official-looking uniforms with the logo of the implementing research institution, carried expensive tablet computers, and – if prompted – carried a copy of a letter from the local government authorizing them to carry out the survey. Thus, to the extent that they influenced the respondent's answer, they should influence it in the direction of expressing pro-government sentiments. On the other hand, social desirability bias from onlookers *deflates* affirmative responses – likely because respondents worry that an onlooker might report back to the NPA. To the extent that self enumeration reduces social desirability bias from both sources, the effects might cancel each other out.

C   Additional Results

None of the respondents who self-enumerated refused to answer the question about high school completion, causing the variable to drop from the models below. The logistic regression models in Columns 2 and 3 return a coefficient for Self Enumeration with virtually no predictive power ($p > .99$) based on the other variables. We report these coefficients for completeness, but they provide virtually no useful information. Separately, we draw readers' attention to the fact that the fixed effects models in Columns 2 and 3 drop nearly 90% of observations. These are barangays in which there was no variation in non-response, i.e. all subjects surveyed in the barangay responded to the question.

Table 3: Non-Response to Placebo Question

	(1)	(2)	(3)
Self		-17.13	-16.14
		(1950.6)	(1417.7)
Forced Choice	0.981**	0.986**	1.083**
	(0.434)	(0.455)	(0.479)
Bystander			0.381
			(0.516)
Self x Bystander			-0.0786
			(2433.8)
Male			-0.146
			(0.497)
Age			-0.0424**
			(0.0172)
Education			-0.425**
			(0.212)
Income			0.0000215
			(0.0000217)
Constant	-5.414***		
	(0.376)		
Barangay FE	No	Yes	Yes
Observations	3028	344	337

Logistic regression with barangay-clustered SE.

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4: Did you graduate high school?

	(1)	(2)	(3)
Self	-0.0208	-0.0625	-0.0431
	(0.0692)	(0.0775)	(0.100)
Bystander			-0.270*
			(0.149)
Self x Bystander			-0.153
			(0.207)
Male			0.0494
			(0.0962)
Age			-0.0518***
			(0.00325)
Income			0.0000782***
			(0.00000811)
Constant	0.0942*		
	(0.0567)		
Barangay FE	No	Yes	Yes
Observations	3044	3017	2963

Logistic regression with barangay-clustered SE.

* $p < .10$, ** $p < .05$, *** $p < .01$

Forced Choice Calculations

For the sensitive item (reporting insurgents), 1,146 subjects answered either "yes" or "no." Probability dictates that about 50% of these answers, or 573, were forced "yeses" from the coin flip, leaving 210 true "yeses" ($782 - 573$), or 36.6% of the 573 total true answers.

For the placebo item (high school completion), 1,432 subjects answered either "yes" or "no." Probability dictates that about 50% of these answers, or 716, were forced "yeses" from the coin flip, leaving 183 true "yeses" ($899 - 716$), or 25.6% of the 716 total true answers.

Timing

Our sensitive and placebo questions came at the end of a survey covering several topics. Average duration was about 26.5 minutes (median 21.3 minutes), excluding seven outliers recorded at greater than 120 minutes.[3]

We observe a small but significant increase in survey duration when using forced choice compared to the other methods. Average duration was 26.2 minutes using direct questioning, 26.4 minutes using self enumeration, and 27.1 minutes using forced choice. While the increase of just under a minute may seem trivial, it accounts for about 67 enumerator hours – or nearly two weeks of employment – over the course of our study. The minimum durations show a similar pattern, at 7.9 minutes for direct questioning, 7.7 minutes for self enumeration, and 8.2 minutes for forced choice. While we do not suggest that these differences should prevent researchers from using randomization devices where they add value, they represent a real, quantifiable cost to doing so. In contrast, the difference in duration between direct and self enumeration is not statistically significant.

Cognitive Load on Enumerators

A common argument against experiment-based survey methods is that they place excessive demands on enumerators. Those giving the survey must understand how the randomized method works, implement it correctly, and explain it to subjects in a way that leads to compliance. In many cases, expecting this level of comprehension and compliance from enumerators is unreasonable. In contrast, self-enumeration is only marginally more complicated for enumerators to implement than direct questioning.

If forced choice depends more heavily on enumerator performance than the other two methods, then we should see greater variation in subjects' responses *across enumerators* under forced choice. To test for this variation, we include enumerator dummies as predictors of willingness to report to the police / claims of completing high school. We also control for onlookers, as we consistently found this to affect responses above, and cluster standard errors by barangay. Models using direct questioning or self enumeration use logistic regression, while those using data generated by forced choice use the "rrlogit" Stata package.

Table 5: Survey Method and Enumerator Effects

	(1) Report Direct	(2) Report Self	(3) Report RR	(4) HS Direct	(5) HS Self	(6) HS RR
Bystander	-0.142	-0.368**	-0.740*	-0.567***	-0.649***	-1.385**

[3]We suspect that in these instances the enumerator forgot to "complete" the survey.

Table 5: Survey Method and Enumerator Effects, Continued

	(0.165)	(0.175)	(0.445)	(0.141)	(0.148)	(0.625)
enumid2	-0.207	0.183	0.614	-0.0872	0.192	-0.743
	(0.455)	(0.494)	(0.800)	(0.312)	(0.441)	(0.960)
enumid3	-0.270	-0.719**	-1.226*	0.240	-0.00792	-0.565
	(0.384)	(0.339)	(0.650)	(0.265)	(0.320)	(0.576)
enumid4	-0.263	0.715**	-0.429	-0.0674	0.283	-1.270*
	(0.344)	(0.337)	(0.563)	(0.244)	(0.320)	(0.663)
enumid5	-0.131	-0.506	-1.152	0.418	0.381	-0.419
	(0.409)	(0.368)	(0.814)	(0.282)	(0.330)	(0.610)
enumid6	0.479	0.182	-0.306	0.0325	-0.118	-1.711**
	(0.353)	(0.305)	(0.512)	(0.233)	(0.282)	(0.716)
enumid7	-0.429	0.348	-4.727	0.152	0.800**	-0.426
	(0.401)	(0.356)	(9.444)	(0.294)	(0.369)	(0.679)
enumid8	-0.249	-0.444	-2.564	-0.0705	0.00729	-13.40***
	(0.417)	(0.478)	(2.105)	(0.382)	(0.454)	(5.028)
enumid9	0.229	-0.260	-0.875	-0.201	-0.298	-0.249
	(0.465)	(0.356)	(0.699)	(0.319)	(0.318)	(0.671)
enumid10	-0.239	-0.165	-3.074*	-0.305	-0.0777	-2.003*
	(0.343)	(0.317)	(1.712)	(0.253)	(0.294)	(1.028)
enumid11	0.427	-0.145	-1.706**	0.842***	0.617**	-0.746
	(0.367)	(0.339)	(0.781)	(0.280)	(0.306)	(0.757)
enumid12	0.248	-0.629	-0.834	0.725*	0.326	-2.286
	(0.809)	(0.481)	(1.411)	(0.401)	(0.384)	(2.231)
enumid13	0.371	0.177	-1.147	-0.125	0.0310	-2.354*
	(0.429)	(0.407)	(0.763)	(0.294)	(0.352)	(1.425)
enumid14	0.0257	-0.489	-1.298**	-0.0260	0.302	-0.597
	(0.352)	(0.361)	(0.621)	(0.245)	(0.313)	(0.584)
enumid15	-0.380	0.0327	-0.304	0.106	-0.0595	-0.928
	(0.368)	(0.355)	(0.549)	(0.305)	(0.339)	(0.659)
enumid16	-0.00905	1.495**	0.911	0.293	0.312	0.248
	(0.639)	(0.602)	(0.938)	(0.445)	(0.388)	(0.843)
Constant	0.517*	0.577**	0.572	0.125	0.0904	0.137
	(0.299)	(0.259)	(0.437)	(0.194)	(0.244)	(0.467)

Table 5: Survey Method and Enumerator Effects, Continued

Observations	1121	1096	1146	1572	1472	1432

Logistic regression with barangay-clustered SE.
* $p < .10$, ** $p < .05$, *** $p < .01$

Table 5 shows suggestive evidence that forced choice is the most susceptible of the methods to differences across enumerators. For the sensitive item (reporting rebels), none of the coefficients on the 15 enumerator dummies are statistically significant; in contrast 3 and 4 coefficients are significant in the self-enumeration and forced choice models, respectively. Using claimed high school completion as a the dependent variable, 2 enumerator dummy coefficients are significant using direct response, 2 using self enumeration, and 5 using forced choice. Recall that because the forced choice method produces population estimates rather than individual observations, it has less statistical power and produces larger standard errors. In other words, all else equal it should be more difficult to find significant results in the RR models. Yet, we see a greater number of enumerator effects compared to the other methods. While these results are not a smoking gun, they are consistent with the conventional wisdom that forced choice is more sensitive to enumerator characteristics and behavior compared to other methods. Considering that the coin-toss method of forced choice is one of the simpler experiment-based survey methods, these results illustrate the need to take enumerator quality and training very seriously when using such methods.

Heterogeneous Effects by Subgroup

It is highly likely that some respondents were more comfortable using tablets than others, and this variation in familiarity could have influenced our results (Bush and Prather 2019). While we do not have a direct measure of familiarity with technology, both age and education are correlated with non-response and answer choice. Tables 6 and 7 below add interaction terms between these variables and survey method to evaluate whether the effects of survey method depend on a respondent's familiarity with technology. We fail to find any

evidence of such a heterogeneous relationship.

Table 6: Heterogeneous Effects on Non-Response

	(1) Sensitive	(2) Sensitive	(3) Placebo	(4) Placebo
Self	-0.207	-0.250*	0	0
	(0.289)	(0.132)	(.)	(.)
Forced Choice	-0.756***	-0.404***	-0.332	1.980**
	(0.293)	(0.133)	(1.049)	(0.984)
Age	-0.00361		-0.0547***	
	(0.00363)		(0.0200)	
Self × Age	0.000775		0	
	(0.00568)		(.)	
Rand. Resp. × Age	0.00663		0.0330	
	(0.00586)		(0.0224)	
Education		0.131***		0.372
		(0.0432)		(0.359)
Self × Education		0.0491		0
		(0.0631)		(.)
Rand. Resp. × Education		-0.0215		-0.584
		(0.0615)		(0.428)
Constant	-0.721***	-1.102***	-3.110***	-6.123***
	(0.188)	(0.0941)	(0.820)	(0.903)
Observations	4502	4494	3028	3023

Logistic regression with barangay-clustered SE.

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 7: Heterogeneous Effects on Willingness to Report

	(1) Sensitive	(2) Sensitive	(3) Placebo	(4) Placebo
Self Enum.	-0.122	0.0814	-0.169	0.126
	(0.305)	(0.133)	(0.291)	(0.742)
Age	-0.0239***		-0.0506***	
	(0.00415)		(0.00401)	
Self Enum × Age	0.00205		0.00283	
	(0.00569)		(0.00582)	
Education		0.259***		4.228***
		(0.0470)		(0.375)
Self Enum. × Education		-0.0658		-0.0927
		(0.0678)		(0.482)
Constant	1.637***	0.0922	2.561***	-5.905***
	(0.219)	(0.0971)	(0.205)	(0.566)
Observations	2217	2211	3044	3038

Logistic regression with barangay-clustered SE.

* $p < .10$, ** $p < .05$, *** $p < .01$