The Program on Governance
and Local Development

# Do List Experiments Run as Expected? Examining Implementation Failure in Kenya, Zambia, and Malawi

Kristen Kao and Ellen Lust

Working Paper
No. 57 2022

The Program on Governance
and Local Development

UNIVERSITY OF
GOTHENBURG

# Do List Experiments Run as Expected? Examining Implementation Failure in Kenya, Zambia, and Malawi

Kristen Kao
University of Gothenburg
kristenkao@gmail.com

Ellen Lust
University of Gothenburg
ellen.lust@gu.se

## Acknowledgments

This publication was supported by the *Swedish Research Council Recruitment Grant* (Swedish Research Council – E0003801), PI: Pam Fredman; *Social Institutions and Governance: Lessons from Sub-Saharan Africa* grant (Swedish Research Council – 2016-01687), PI: Ellen Lust; *The Governance Challenge of Urbanization* grant (FORMAS – 2016-00228), PI: Ellen Lust. The authors would like to thank reviewers from IJPOR and members of the GEPOP group for their comments. We would also like to thank Cecilia Ahsan Jansson and Erica Ann Metheney for assisting with data compilation and Rose Shaber-Twedt for editing assistance.

## Abstract

A fundamental premise of list experiments is that they allow respondents to hide sensitive attitudes and behaviors among a list of other items. List experiments accomplish this by asking the respondent to count both sensitive and innocuous items, rather than answering questions directly about each item. Social scientists widely employ list experiments to overcome sensitivity bias but have not yet systematically studied whether the complicated nature of these experiments leads to implementation errors. Analysis of a list experiment across three countries suggests that respondents reveal their direct responses to list items more than half of the time. This problem is particularly prevalent among less educated and older respondents, and the complicated nature of the question mediates the relationships between age and education and revealing answers. We encourage scholars to include questions about implementation problems as standard follow-ups in list experiments to understand if they worked properly in the field.

*Keywords*: List experiment, survey methodology, Sub-Saharan Africa, sensitivity bias, social desirability bias

## 1. Introduction

In the past three decades, social scientists have increasingly employed list experiments to study sensitive issues such as prejudice (Hatchett & Schuman, 1975) and voting buying (Gonzales-Ocantos et al., 2012; Çarkoğlu & Aytaç, 2015). Such attitudes and behaviors are often illegal or viewed as immoral; therefore, survey respondents are likely to falsely answer questions with a non-sensitive response. List experiments aim to overcome sensitivity bias by allowing respondents to "hide" truthful responses to a sensitive item among their responses to a list of control items (Bradburn et al., 1978; DeMaio, 1984; Kuklinski et al., 1997). They do so by asking the respondent to report only the count of items they agree with or have engaged in, rather than directly answering questions about each item in the list. Unlike direct questioning, no one listening to the interview can identify if the respondent agreed with or engaged in the sensitive item.

Enthusiastic faith in list experiments – to the extent they are dubbed "statistical truth serum" (Glynn, 2013) – raises questions about the drawbacks and accuracy of the technique. Scholars have focused on the technical aspects of the method, criticizing it as inefficient and requiring a high number of respondents to detect effects (Corstange, 2009; Blair et al., 2020). Furthermore, there are multiple important conditions list experiments must satisfy to be accurate (Blair & Imai, 2012; Glynn, 2013).

Scholars also speculate that there is significant potential for surveyors or respondents to botch the implementation of list experiments. List experiments have a long and confusing format (e.g., Kramon & Weghorst, 2019, p. 237), and decades of survey research find that complicated survey questions are difficult to answer, particularly for certain subgroups (Krosnick, 1991). Yet, there is no systematic study of just how prevalent implementation problems are for this "statistical truth serum" or whether problems are more common among some respondent groups that are often of interest in such studies.

In this study, we examine the extent to which list experiments suffer from implementation failure and lead respondents to reveal answers, despite being asked not to do so. We expect respondents often fail to understand list experiments, casting doubt on their ability to procure more truthful responses than direct questioning. Moreover, we suspect this problem is more prevalent among the less educated and elderly. Thus, outcomes from these experiments are likely less precise than is often implied, and apparent differences in sensitivity of subjects among subpopulations may instead reflect differences in implementation.

## 2. Expectations of Implementation Errors

We present hypotheses for two possible implementation error sources: 1) *disclosure of responses,* e.g., answers to specific list items are revealed, and 2) *cognitive overload,* leading respondents to be unable or unwilling to consider each item carefully. Response disclosure is our primary outcome of interest, as it completely undermines the purpose of a list experiment. However, we expect cognitive overload may help explain why implementation errors occur, and that the error is related to the respondent's faculties. All hypotheses were pre-registered with the Open Science Framework public registry.[1]

To test these hypotheses, beyond the traditional count outcome to the list experiment, we asked three additional binary questions that comprise our outcome variables in this study. The first probed the respondent: "Do you feel that the last question was too complicated to answer properly?" The next two questions were posed to the enumerator only: 1) (DO NOT READ:)[2] Respondent was able to count up the number of items without my help, and 2) (DO NOT READ:) Respondent revealed to me which items they had done in the past. (For exact question wording, see Appendix I). The first and second questions indicate *cognitive overload*, whereas the last question captures *disclosure of responses.*

Our first set of hypotheses concerns the prevalence of implementation errors. For this set of hypotheses, we selected 40 percent as a high proportion of respondents, as there is empirical precedent for this expectation (Kramon & Weghorst, 2019).[3] Arguably though, even a 5 or 10 percent error would be enough to cast doubt on most list experiment outcomes (See Blair et al.'s (2020) analysis of list experiments conducted before 2017)[4]. Specifically, we expected:

> *H1a-c.* High proportions of surveyors (over 40 percent) report respondents a) revealed their specific responses to the items in the list; b) were unable to count their affirmative answers to the list items without help; c) reported they found the list experiment too complicated to answer.

[1] The numbering of hypotheses and order of presentation was different in the preregistration. This has been revised to facilitate presentation. Exact wording is found in Appendix V.
[2] Enumerators were instructed that DO NOT READ meant that they should not read the question aloud to the respondent.
[3] Kramon and Weghorst's (2019) measurement strategy is more indirect than ours. They compare the counts of affirmative responses to direct questions with counts from the list experiment, noting the proportion of respondents whose counts deviate from 0 as those for whom there was a "list experiment failure."
[4] Their study included 487 distinct experiments published in 154 papers.

Moreover, we expected less educated and older respondents to find list experiments more difficult than others. They are thus more likely to make implementation errors, violating the assumptions of the experimental design.[5] Here, there is precedence for finding that the less educated are more prone to list experiment implementation errors (Kramon & Weghorst, 2019). Specifically:

> *H2a-c.* Less-educated interviewees (primary education or below) are more likely to have difficulty responding to list experiments than more educated interviewees. Thus, interviewers are more likely to report that these respondents a) reveal their answers; b) cannot count their answers without help; c) are more likely to report that they find the list experiment too complicated to answer.

> *H3a-c.* Elderly respondents (age 60 and above) are more likely to have difficulty responding to a list experiment than younger respondents. Therefore, interviewers are more likely to report older respondents a) reveal their answers; b) cannot count their answers without help; c) are more likely to report that they find the experiment too complicated to answer.

> *H4*. For both subpopulations, reporting of the list experiment as too complicated, and help was needed to count items mediate the primary outcome of *disclosure of responses*.

Finally, we suspected that confusion might particularly adversely affect the experimental treatment group. Because the treatment item is sensitive, respondents are more likely to feel anxious about answering the question. Treatment group respondents also deal with one more item than control group respondents, increasing their cognitive burden. We note that this is our only truly experimental outcome in this study, as it is the only analysis that will use randomization inference. Specifically:

> *H5a-c.* Interviewers are more likely to report that treatment group respondents a) reveal their answers; b) cannot count their answers without help; c) are likely to find responding to the list experiment more difficult than those in the control group.

> *H6*. Reporting of the list experiment as too complicated and that help was needed to count items mediate our primary outcome of *disclosure of responses* among the elderly and the less educated.

[5] This study does not present results related to respondent age.

In addition, we explore our outcome responses by ethnicity, gender, and wealth. However, we expected failures in list experiment implementation to arise from differential skills and faculties, not demographics.

## 3. Method

The experiment was included within a larger survey on governance and local development in Kenya, Malawi, and Zambia. This survey was conducted in five regions across the three countries. Enumerators were well-trained in the experimental protocols; the researchers spent months in the field before the survey, got to know the local teams well, and conducted in-depth training with clear instructions to ensure proper implementation of the list experiment. Treatment and control groups were randomized by programming embedded within the SurveytoGo platform and were thus independent of the interviewer and researchers.

### 3.1 The List Experiment

The list experiment was designed to study vote-buying, an issue for which list experiments have been employed with notable findings. For instance, Gonzales-Ocantos et al. (2012) found that, when asked directly, only 2 percent of Nicaraguan voters admitted to being offered gifts or services in exchange for their votes, compared to nearly one in four in a list experiment. Çarkoğlu and Aytaç's (2015) list experiment in Turkey found that 35 percent of the sample received vote-buying offers, while only 16 percent reported this behavior in a direct question. Blair et al. (2020) considered 19 list experiments on vote-buying and found the average reporting error is –8 points with a 95% confidence interval ranging between –13 and –3 points.

The experimental design took into consideration best practices in list experiment methodology. Items were selected to be straightforward and easy to recall.[6] To avoid ceiling effects, we included two items with a low prevalence of affirmative responses from a survey conducted in Malawi in 2016 (n=8,000). Moving to another village received 29 percent of positive responses among the 2016 sample, and having water piped into one's dwelling was affirmed by 2 percent of respondents (with Afrobarometer (2018) reporting 8, 3, and 11 percent for Kenya, Malawi, and Zambia respectively on the latter

[6] Ceiling effects occur when treated respondents tend to answer the maximum number of items on the list (Glynn 2013), whereas floor effects refer to respondents in the treatment group answering zero items (Blair and Imai 2012). Both effects inhibit proper analysis of list experiments.

question). On the other hand, only 2 percent of respondents denied visiting others in their neighborhood or village. The sensitive item asked respondents if they have ever "become more likely to vote for a candidate because they offered/gave you gifts, money, or personal favors." We also included cartoons for respondents to increase precision (Kramon & Weghorst, 2019). The item order was randomized to avoid recency effects (Krosnick & Alwin, 1987). Finally, we wrote out clear instructions for enumerators, including asking enumerators to turn 180 degrees away from the respondent while they were counting the items. We did this upon the advice of others (Kramon & Weghorst, 2019)[7] to ensure the respondent felt they had ample time, ability, and privacy to count the items.

The experimental prompt read as follows, with unvocalized enumerator instructions in italics and parentheses: "Please listen to this list carefully and tell me how many of these things you have done in the past. DO NOT TELL ME WHICH THINGS you have done in the past, ONLY THE TOTAL NUMBER of things. I will turn around to read this list to allow you to count and answer. (*Enumerator: Turn around and continue reading SLOWLY*)

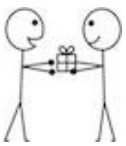
Please count, have you ever:

Moved to a different village or neighborhood than this one.

Had water piped into your current dwelling

Visited with others in this village or neighborhood

Become more likely to vote for a candidate because they offered/gave you gifts, money, or personal favors



[7] The authors thank David Nickerson for this helpful advice through personal correspondence (February 19, 2019).

(*Enumerator: Ask the respondent if they would like you to read the list again. If yes, read it again; if not, continue reading.*) How many of these things have you done in the past? 0, 1, 2, 3, or 4?[8] I am going to hand you the tablet now. Please press on the circle next to the number of things you have done in the past and tell me once you have done so."

The enumerator turned around when the experiment was read out in case the respondent wanted to count the items on their fingers or employ an otherwise public means to track the items in which they had engaged. Then, the enumerator reminded the respondent that they should only reveal the number, not the items in which they had engaged. The enumerator then pointed to each picture and list item, explaining each one again and the answer options. The enumerator then turned around again to allow the respondent to count based on the pictures and select an answer on the tablet. The procedure was designed to give the respondent a sense of utmost privacy when responding, thereby encouraging truthful responses. If anything, this procedure should have worked against us finding implementation errors in our list experiment.

### 3.2 Measures

The three follow-up questions mentioned above provided our three primary dependent variables: *Complicated, Counting Help,* and *Disclosure*. *Complicated* was the binary variable created from the question asking whether the respondent viewed the question as too complicated to answer properly. *Counting Help* was the binary variable created from the interviewer's response to whether the respondent could count the number of items without help. If the respondent could not provide an answer to the outcome question, the enumerator was instructed to provide further guidance on counting. For instance, the respondent could hold up a finger for each item in the list she or he had engaged in previously. Since the enumerator's back is turned during the reading of the list items, the respondent can still prevent disclosure of engagement in specific items using this method. Finally, *Disclosure* was the binary variable stemming from the enumerator's response to whether the respondent revealed which items they had done in the past. *Complicated* takes the value 1 if the respondent felt that it was too complicated to answer properly and 0 otherwise. *Counting Help* takes the value 1 if the respondent

[8] The control group saw options of 0-3.

needed help counting the number of items and 0 otherwise. *Disclosure* takes the value 1 if the respondent revealed having done any of the list items to the enumerator and 0 otherwise.[9]

We consider four primary independent variables. *Elderly* is a binary variable taking the value 1 if the respondent was at least 60 years old and 0 if not. *Primary or Less* is a binary variable taking the value 1 if the respondent had no, only informal, or only primary schooling and 0 if the respondent had more schooling. *Insufficient Income* takes the value 1 if the respondent reported having difficulties/great difficulties covering their expenses and 0 if the respondent could cover needs/could save money. Finally, *Treatment* takes the value of 1 if the respondent was randomly assigned to the treatment group, receiving the fourth list item: "Become more likely to vote for a candidate because they offered/gave you gifts, money, or personal favors" and 0 if not.

## 4. Analyses and Results

We assess our hypotheses in three ways. First, we examine frequencies to determine how likely respondents are to: reveal responses to list items, find the experiment too complicated, or require help answering the question. Second, we fit OLS regression and binary logit models, using our outcome questions as dependent variables and treatment indicators as independent variables. Finally, we employ mediation analysis to consider whether respondents finding the experiment too complicated or being unable to count their answers without help mediate the effects of subpopulation indicators on revealing responses to specific list items.

We begin by examining the descriptive hypotheses (H1a-c). Enumerators reported significant problems: 55 percent stated their respondents disclosed responses to specific items on the list, and 79 percent could not count their answers without help (thus supporting H1a and H1b). However, only 22 percent of respondents declared the questions too difficult to answer, thus failing to support H1c. The Pearson Correlation between *Disclosure* and *Complicated* is 0.0257, between *Complicated* and *Counting Help* is 0.2456, and between *Counting Help* and *Disclosure* is -0.1175. We suspect the negative correlation stems from the fact that those who ask for help in counting their answers are more likely to answer the question with the number of items, avoiding disclosure.

[9] Only the first question to the respondent allowed for non-response – this was only about 3 percent of the sample. For the analyses using this question, respondents answering do not know or refuse to answer were dropped out.

Logistic regression analysis was employed to examine the relationship between education, age, treatment, and our three dependent variables. We find evidence supporting H2a-c and H3a-c – that less-educated and elderly respondents, respectively, find list experiments more challenging. However, we lack support for H5a-c – that those in the treatment group are more likely to reveal their answers, need help counting, or find the experiment too complicated (See Table 1 for full results).

Standard controls for demographic variations in all survey research were employed. We found a small effect (0.18) of gender on a respondent's admission that the experiment was too complicated ($p<0.05$) and no significant effect of wealth on any of the three outcomes. This reinforces our expectation that problems associated with list experiments are due to differences in respondents' abilities to handle such questions, not simply demographics. We also included fixed effects for ethnic groups, as some groups may be targeted for vote-buying.

**Table 1: Logistic Regression of Complicated, Counting Help and Disclosure**

| | (1) Complicated | (2) Counting Help | (3) Disclosure |
|---|---|---|---|
| Treatment | 0.122 | -0.0262 | 0.0370 |
| | (0.0850) | (0.0775) | (0.0711) |
| Elderly | 0.550*** | 0.600*** | 0.225^ |
| | (0.137) | (0.127) | (0.128) |
| Primary or Less | 0.290** | 0.556*** | 0.265*** |
| | (0.0974) | (0.0885) | (0.0824) |
| Insufficient Income | 0.119 | 0.109 | -0.101 |
| | (0.104) | (0.0952) | (0.0858) |
| Female | 0.180* | 0.0737 | -0.000939 |
| | (0.0890) | (0.0806) | (0.0734) |
| Lusaka | 0.952*** | -0.478* | -0.804*** |
| | (0.244) | (0.208) | (0.198) |
| Malawi Border | 0.514* | -0.261 | 0.167 |
| | (0.210) | (0.163) | (0.162) |
| Nairobi | 0.383 | -0.375 | 0.101 |
| | (0.675) | (0.610) | (0.562) |
| Zambia Border | 0.857*** | -0.502** | -0.691*** |
| | (0.228) | (0.187) | (0.182) |
| Constant | -1.960*** | -0.848*** | 0.237 |
| | (0.309) | (0.276) | (0.259) |
| Ethnic Fixed Effects | Yes | Yes | Yes |
| Observations | 3,323 | 3,403 | 3,429 |

Robust standard errors are in parentheses *** $p < .001$, ** $p < .01$, * $p < .05$, ^ $p < .10$.

We use the R Package for Causal Mediation, based on Tingley et al. (2014), to examine the extent to which *Counting Help* and *Complicated* mediate the relationships between age or less education and a respondent's likelihood to disclose their responses – our primary outcome of interest (H4 and H6). We examined four mediations. The first two concerned the relationship between *Elderly* and *Disclosure,* considering whether it was mediated by the extent to which 1) the question was *Complicated* or 2) the elderly needed help counting. The second two concerned the relationship between the less-educated and *Disclosure,* considering whether it was mediated by the extent to which 1) the question was *Complicated* or 2) the less-educated needed help counting. Results are summarized in Table 2 (See Appendix IV for full results).

**Table 2: Average Proportion of Variance in Disclosure Mediated**

| **Independent Variable** | **Mediator** | **Estimate** | ***p*-value** |
|---|---|---|---|
| Elderly | *Complicated* | 0.078 | 0.0564 |
| Elderly | *Counting Help* | 0.041 | 0.830 |
| Primary or Less | *Complicated* | 0.036 | 0.0296 |
| Primary or Less | *Counting Help* | 0.043 | 0.4664 |

We find some evidence that *Complicated* mediates the relationship between our findings among the elderly and the less educated and *Disclosure*. The mediation analysis suggests that approximately 8 percent of the variance in *Disclosure* among the elderly is explained by reports of the list experiment being *Complicated,* although the estimate is only marginally statistically significant ($p<0.10$). Approximately 4 percent of the variance ($p<0.05$) of *Disclosure* among respondents with a *Primary or Less* education level is mediated by reports that the experimental instructions were too *Complicated*. There is no evidence that requiring help to count mediates the relationship between elderly or less-educated respondents and *Disclosure*.

## 5. Conclusion

List experiments have been widely employed to overcome problems of sensitivity bias in survey research. Their value, however, rests on the assumption that implementation does not lead to response disclosure and that, at the very least, disclosures are not systematic across certain demographic subpopulations.

Our study draws these assumptions into question. We find that respondents disclosed their answers more than half (55 percent) of the time and that the vast majority (79 percent) required help counting the answers. Moreover, we provide evidence that the less educated and older respondents are more likely to experience difficulties answering the questions. They are also more likely to disclose their answers, at least in part, because they find the experiment complicated. Although we do not claim that all list experiments suffer from implementation errors to the extent that those in our experiment seem to, our findings indicate the need for further investigation of these issues.

Recognizing the implementation problems of list experiments is important. A census of list experiments conducted between 1984 and 2017 led scholars to conclude that sensitivity bias is actually rather limited (Blair et al., 2020, p. 1298). However, our results suggest that this finding may reflect the prevalence of implementation errors more than reality. Apparent differences in attitudes or reported behaviors may result from different abilities to engage with the experiment. We encourage scholars to follow advice on reducing the complexity of list experiments (see, e.g., Kramon & Weghorst, 2019) and include questions about implementation problems as standard follow-ups to understand if they worked properly in the field. Doing so will help scholars track and potentially account for subgroup differences in faculties leading to list experiment implementation errors.

## References


Afrobarometer Data. (2018). *Malawi, Kenya, and Zambia.* Round 7. Available at: http://www.afrobarometer.org.

Blair, G., Coppock, A. & Moor, M. (2020). "When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments," *American Political Science Review,* 114(4), pp. 1297-315.

Blair, G. & Imai, K. (2012). "Statistical analysis of list experiments," *Political Analysis,* 20(1), pp. 47-77.

Çarkoğlu, A. & Aytaç, E. (2015). "Who gets targeted for vote-buying? Evidence from an augmented list experiment in Turkey," *European Political Science Review,* 7(4), pp. 547-66.

Corstange, D. (2009). "Sensitive questions, truthful answers? Modeling the list experiment with LISTIT," *Political Analysis,* 17(1), pp. 45-63.

DeMaio, T. J. (1984). "Social desirability and survey measurement: A review," in Turner, C. F. & Martin, E. (eds.) *Surveying Subjective Phenomena Vol. 2.* New York: Russell Sage Foundation, pp. 257-82.

Glynn, A. (2013). "What can we learn with statistical truth serum? Design and analysis of the list experiment," *Public Opinion Quarterly,* 77(1), pp. 159-72.

Gonzales-Ocantos, E., Kiewiet de Jong, C., Meléndez, C., Osorio, J. & Nickerson, D. (2012). "Vote buying and social desirability bias: Experimental evidence from Nicaragua," *American Journal of Political Science,* 56(1), pp. 202-17.

Hatchett, S. & Schuman, H. (1975). "White respondents and race-of-interviewer effects," *Public Opinion Quarterly,* 39(4), pp. 523–8.

Kramon, E. & Weghorst, K. (2019). "(Mis)measuring sensitive attitudes with the list experiment: Solutions to list experiment breakdown in Kenya," *Public Opinion Quarterly,* 83(S1), pp. 236-63.

Krosnick, J.A. & Alwin, D. F. (1987). "An evaluation of a cognitive theory of response-order effects in survey measurement," *Public Opinion Quarterly,* 51(2), pp. 201-19.

Krosnick, J. A. (1991). "Response strategies for coping with the cognitive demands of attitude measures in surveys," *Journal of Cognitive Psychology,* 5(3), pp. 213–36.

Kuklinski, J. H., Cobb, M. D. & Gilens, M. (1997). "Racial attitudes and the 'New South,'" *Journal of Politics,* 59(2), pp. 323– 49.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. (2014). "Mediation: R Package for causal mediation analysis," *Journal of Statistical Software*, 59(5), pp. 1–38.

# Appendices

## Appendix I. Enumerator Protocols List Experiment

- You will be reading a list of either 3 or 4 items to a respondent, clearly and slowly. You will ask the respondent how many of these he or she has done in the past.
- You DO NOT want to know **which** items exactly on the list the respondent has actually done.
- Instead, you just want to know HOW MANY of these items the respondent has done.
- The whole point is to keep the items the respondent has done confidential, but still allow you to record the correct number of these items.
- The items on the list will randomize order every time the survey is run and we want you to read the items in the order in which they appear on the screen.
- You will first read the prompt, then you will turn around as you read the list of items. Once you finish, you will face the respondent again and hand him/her the tablet. Please point to the pictures and say the list items again while reminding the respondent that you do not want to know which items he or she has done. Point to the answer options and read them out for the respondent as well, showing him/her where to press. Then turn back around to allow the respondent anonymity as he/she records the number of times.

## Appendix II. Summary Statistics

Table 1. Distributions of Gender

| Country | Male | Female |
|---|---|---|
| Kenya | 45% (462) | 55% (562) |
| Malawi | 36% (437) | 64% (784) |
| Zambia | 41% (502) | 59% (734) |
| Total | 1401 | 2080 |

Table 2. Distributions of Educational Attainment

| Country | None or little education | Primary schooling | Secondary schooling | Post-Secondary schooling |
|---|---|---|---|---|
| Kenya | 1% (12) | 23% (236) | 46% (473) | 29% (299) |
| Malawi | 12% (145) | 58% (708) | 27% (323) | 3% (40) |
| Zambia | 9% (111) | 46% (566) | 36% (445) | 9% (112) |
| Total | 268 | 1510 | 1241 | 451 |

Table 3. Distributions of Wealth

| Country | Can afford costs | Insufficient income to cover needs |
|---|---|---|
| Kenya | 48% (484) | 52% (527) |
| Malawi | 14% (168) | 86% (1044) |
| Zambia | 27% (328) | 73% (897) |
| Total | 980 | 2468 |

## Appendix III. Distributions of Variables

Table 4. Distributions of *Group*

| | Control | Treatment |
|---|---|---|
| n | 1738 | 1743 |
| % | 49.93 | 50.07 |

Table 5. Distributions of *Disclosure*

| | No | Yes |
|---|---|---|
| n | 1567 | 1914 |
| % | 45.02 | 54.98 |

Table 6. Distributions of *Counting Help*

| | No | Yes |
|---|---|---|
| n | 1031 | 2450 |
| % | 29.62 | 70.38 |

Table 7. Distributions of *Complicated*

| | No | Yes | DK | RA |
|---|---|---|---|---|
| n | 2614 | 778 | 61 | 28 |
| % | 75.09 | 22.35 | 1.75 | 0.80 |

Table 8. OLS Linear Probability Model (Outcomes: *Complicated*, *Counting Help*, *Disclosure*)

| | (1) | (2) | (3) |
|---|---|---|---|
| | Complicated | Counting Help | Disclosure |
| Treatment | 0.0202 | -0.00580 | 0.00858 |
| | (0.0144) | (0.0154) | (0.0167) |
| Elderly | 0.108*** | 0.135*** | 0.0523^ |
| | (0.0291) | (0.0297) | (0.0295) |
| Primary Or Less | 0.0485** | 0.111*** | 0.0619*** |
| | (0.0164) | (0.0175) | (0.0193) |
| Insufficient Income | 0.0196 | 0.0200 | -0.0235 |
| | (0.0168) | (0.0176) | (0.0201) |
| Female | 0.0304* | 0.0149 | -4.83e-05 |
| | (0.0148) | (0.0158) | (0.0173) |
| Lusaka | 0.160*** | -0.107* | -0.194*** |
| | (0.0395) | (0.0453) | (0.0469) |
| Malawi Border | 0.0795** | -0.0617 | 0.0387 |
| | (0.0301) | (0.0384) | (0.0387) |
| Nairobi | 0.0574 | -0.0839 | 0.0244 |
| | (0.108) | (0.113) | (0.132) |
| Zambia Border | 0.143*** | -0.112** | -0.169*** |
| | (0.0358) | (0.0424) | (0.0438) |
| Constant | 0.127* | 0.317*** | 0.557*** |
| | (0.0559) | (0.0580) | (0.0618) |
| Ethnic Fixed Effects | Yes | Yes | Yes |
| Observations | 3,357 | 3,442 | 3,442 |
| R-squared | 0.048 | 0.059 | 0.059 |

Robust standard errors in parentheses *** p<0.001, ** p<0.01, * p<0.05, ^ p<0.10

Table 9. Pearson's Correlation Coefficient for *Disclosure*, *Complicated* and *Counting Help*

| | Disclosure | Complicated | Counting Help |
|---|---|---|---|
| Disclosure | 1 | | |
| Complicated | 0.0257 | 1 | |
| Counting Help | -0.1175 | 0.2456 | 1 |

## Appendix IV. Mediation Analysis

Table 10: Set Up of Mediation Analysis of Complicated on *Disclosure* Among the Elderly

| **Outcome Variable** | Disclosure |
|---|---|
| **Main Independent Variable** | Elderly |
| **Mediating Variable** | Complicated |
| **Model Type** | Logistic |

Table 11: Mediation Analysis of Complicated on *Disclosure* Among the Elderly

| | Estimate | 95% CI Lower | 95% CI Upper | p-value |
|---|---|---|---|---|
| ACME (control) | 0.0050825 | 0.0008871 | 0.0107392 | 0.0164 |
| ACME (treated) | 0.0048859 | 0.0008520 | 0.0103005 | 0.0164 |
| ADE (control) | 0.0548751 | -0.0044773 | 0.1118510 | 0.0692 |
| ADE (treated) | 0.0546785 | -0.0044728 | 0.1113416 | 0.0692 |
| Total Effect | 0.0597610 | -0.0001463 | 0.1155279 | 0.0504 |
| Prop. Mediated (control) | 0.0796192 | -0.0736741 | 0.5299986 | 0.0668 |
| Prop. Mediated (treated) | 0.0763012 | -0.0714440 | 0.5288856 | 0.0668 |
| ACME (average) | 0.0049842 | 0.0008705 | 0.0105455 | 0.0164 |
| ADE (average) | 0.0547768 | -0.0044750 | 0.1116163 | 0.0692 |
| Prop. Mediated (average) | 0.0779602 | -0.0730045 | 0.5295626 | 0.0668 |

Table 12: Set Up of Mediation Analysis of Complicated on *Disclosure* Among Those with Primary or Less Education

| **Outcome Variable** | Disclosure |
|---|---|
| **Main Independent Variable** | Primary Or Less |
| **Mediating Variable** | Complicated |
| **Model Type** | Logistic |

Table 13: Mediation Analysis of Complicated on *Disclosure* Among the Elderly

| | Estimate | 95% CI Lower | 95% CI Upper | p-value |
|---|---|---|---|---|
| ACME (control) | 0.0024176 | 0.0002383 | 0.0056222 | 0.0212 |
| ACME (treated) | 0.0023539 | 0.0002326 | 0.0054666 | 0.0212 |
| ADE (control) | 0.0593521 | 0.0231864 | 0.0971146 | 0.0016 |
| ADE (treated) | 0.0592884 | 0.0231509 | 0.0970440 | 0.0016 |
| Total Effect | 0.0617060 | 0.0254969 | 0.0993776 | 0.0008 |
| Prop. Mediated (control) | 0.0366391 | 0.0038929 | 0.1264741 | 0.0220 |
| Prop. Mediated (treated) | 0.0356451 | 0.0037008 | 0.1245483 | 0.0220 |
| ACME (average) | 0.0023858 | 0.0002341 | 0.0055493 | 0.0212 |
| ADE (average) | 0.0593202 | 0.0231696 | 0.0971029 | 0.0016 |
| Prop. Mediated (average) | 0.0361421 | 0.0037732 | 0.1253366 | 0.0220 |

Table 14: Logistic Regression Results Mediation Analysis of *Complicated* on *Disclosure* Among the Elderly and Primary Education or Less

| | *Dependent variable:* | |
|---|---|---|
| | *Complicated* | *Disclosure* |
| Complicated | | 0.219* |
| | | (0.088) |
| Elderly | 0.545*** | 0.240$^{+}$ |
| | (0.137) | (0.133) |
| Insufficient Income | 0.118 | −0.117 |
| | (0.105) | (0.087) |
| Primary Or Less | 0.287** | 0.261** |
| | (0.098) | (0.084) |
| Gender (male) | −0.178* | 0.007 |
| | (0.089) | (0.075) |
| Region Lusaka | 0.953*** | −0.819*** |
| | (0.244) | (0.200) |
| Region Malawi Border | 0.516* | 0.186 |
| | (0.211) | (0.163) |
| Region Nairobi | 0.396 | 0.162 |
| | (0.659) | (0.541) |
| Region Zambia Border | 0.858*** | −0.714*** |
| | (0.230) | (0.184) |
| Constant | −1.717*** | 0.284 |
| | (0.298) | (0.256) |
| Observations | 3,357 | 3,357 |
| Akaike Inf. Crit. | 3,569.674 | 4,524.857 |

*Note:* + p<0.1; * p<0.05; ** p<0.01; *** p<0.001

## Appendix V. Original Pre-Analysis Plan

Do List Experiments Run as Expected? An Examination of Implementation Failure in Kenya, Zambia, and Malawi

**Motivation**

Vote buying – the exchange of particularistic goods and services for political support (Kitschelt and Wilkinson 2007; Nichter 2014) – is said to dominate the politics of the global south. Yet, it remains an elusive concept for scholars to study due to what is known as social desirability bias (Bradburn et al. 1978; DeMaio 1984). Such bias occurs when survey participants falsely answer a question in order to provide a more socially desirable response. Such bias can lead to either underreporting or overreporting of attitudes or behaviors such as prejudice (Hatchett and Schuman 1975) or voting in elections (Campbell 1960). Engaging in vote buying (either as the buyer or seller) is illegal in most countries (Schaffer 2007) and may be considered to be immoral or detrimental to society (Gonzales-Ocantos et al. 2014); thus, survey questions about its practice are often expected to be plagued by social desirability bias. Researchers may also include a list experiment in their research in order to demonstrate that responses to their questions of interest are *not* sensitive. For example, Kao and Revkin (Forthcoming) use such an experiment in Mosul, Iraq to show that the estimated sensitivity bias to a direct question about the Islamic State's governance performance could not be distinguished from zero.

To solve the issue of social desirability bias, social scientists often employ list experiments, a statistical technique that has been in use for more than 3 decades (Miller 1984). List experiments give respondents the opportunity to truthfully report their behavior while remaining anonymous to the interviewer (Kuklinski et al. 1997). Essentially, respondents can "hide" truthful responses to a sensitive item among responses to a list of control questions. Instead of answering a sensitive question directly, the respondent is read a set of binary questions; she counts up and reports how many affirmative responses to this list of questions are applicable to her. Respondents are randomly selected to receive either a list of innocuous items (control group) or the same list including the sensitive item (treatment group). A comparison of the average count of affirmative items among those in the control with those in the treatment group allows the researcher to reveal the true prevalence of the sensitive item. If the difference is not statistically significant, the item can be considered to be non-sensitive or at least to the extent that it would greatly affect average responses from among a sample.

Concerning vote buying specifically, list experiments have been widely employed.[10] Gonzales-Orcantes et al. (2012) find that when asked directly, only 2 percent of Nicaraguan voters admit to being offered gifts or services in exchange for their votes compared to nearly one in four reporting vote buying offers in a list experiment. Likewise, Çarkoğlu and Aytaç (2015)'s list experiment in Turkey finds that 35 percent of the sample received vote buying offers, while only 16 percent reported this behavior in a direct question. Blair, Coppock, and Moore (2020, 1309) consider 19 studies that have employed a list experiment to study vote buying and find that the average reporting error is –8 points with a 95% confidence interval ranging between –13 and –3 points.

[10] See Blair, Coppock, and Moore (2020) for a list of 19 studies that have employed a list experiment to study vote buying.

Such enthusiastic faith in the list experiment to reveal sensitivity bias – to the extent that they are dubbed the "statistical truth serum" (Glynn 2013) – motivates questions about the drawbacks of the technique and how accurate it really is. One major criticism of the method is that it is inefficient, requiring a high number of respondents to detect effects (Corstange 2009; Blair, Coppock, and Moore 2020). The literature also notes important conditions that list experiments must satisfy in order to be accurate including consideration of "ceiling" and "floor" effects (Blair and Imai 2012; Glynn 2013) and assumptions that inclusion of the sensitive item does not affect responses to the control items (Blair and Imai 2012). In this study we focus on the extent to which they work as expected in practice. Based on decades of survey research demonstrating that complicated survey questions are difficult to answer and particularly so for subgroups of respondents (Krosnick 1991), we have good reason to expect that outcomes from these experiments are more imprecise than is often implied in research that employs them.

Scholars speculate that there is high potential for either surveyors or respondents to botch the implementation of list experiments due to their long and confusing format (e.g., Kramon and Weghorst 2019), but there is little systematic, empirical evidence of just how prevalent problems are in the field. In particular, we expect that list experiments often suffer a major implementation failure during fielding: respondents end up revealing their answers to list items despite being asked not to do so. As far as we know, the extent of list experiment implementation failure has not been studied before. As noted above, we have reasons to believe that respondents often fail to understand list experiments, including the purpose behind them meaning they do not realize the anonymity this experiment affords them. This is a serious concern as it invalidates the main purpose of the experiment. We also suspect that these implementation errors occur more frequently among certain subpopulations. Thus, using list experiments to detect sensitivity of subjects among subpopulations may instead reveal differences in implementation. Finally, since this problem has not been directly studied before, scholars do not fully understand the reasons these errors in list experiments may occur.

Our research seeks to fill these gaps. We interrogate two possible reasons for why implementation errors may occur: 1) *violations of anonymity,* such that answers to specific list items are revealed, and 2) *cognitive overload,* leading respondents to be unable or unwilling to carefully consider each item and count up their affirmative answers to all of the list items in order to report them at the end of the question. Both of these implementation errors invalidate the purposes of a list experiment and the latter error is expected to explain why the former occurs.

The study will make a contribution to the field regardless of the outcome. It will either help to validate the reliability of list experiments or demonstrate the extent to which the design is prone to implementation error, particularly among populations in the global south and among uneducated populations in particular. Further, the study outcomes will shed some insight into why these errors, if present, may occur. Doing so, it aims to help researchers employ the technique appropriately and identify solutions for improving the implementation of list experiments in the field.

**Hypotheses**

First, we develop generalized hypotheses using non-experimental behavior. We suspect that high proportions of respondents end up revealing their answers to some or all of the specific list items to enumerators. This violation of anonymity is our primary outcome of interest as it completely undermines the purpose of a list experiment. For the hypotheses below, we selected 40 percent as a high proportion of respondents since we believe this to constitute extremely significant errors in

reporting. There is also empirical precedent for this expectation: Kramon and Weghorst (2019) demonstrate that more than 40 percent of their sample of Kenyans were prone to reporting errors. One could argue, however, that as a little as 5 or 10 percent error would be enough to cast doubt on most list experiment outcomes, as Blair, Coppock, and Moor demonstrate in their analysis of a census of list experiments conducted before 2017.[11]

We also probe why such errors are likely to occur by asking about whether the experiment was too complicated to understand and whether counting up the list items was difficult without help. These two outcomes are secondary, serving more as mechanisms to why our primary outcome occurs. Given our large sample size, within our control group we will also compare the counts of affirmative responses to direct questions with counts from the list experiment, noting the proportion of respondents whose counts deviate from 0 as those for whom there was a "list experiment failure" (Kramon and Weghorst 2019, 237).

> H1. High proportions of surveyors (over 40 percent) report that respondents revealed their specific responses to the items in the list.
>
> H2. High proportions of respondents (over 40 percent) report that they find the list experiment too complicated to answer.
>
> H3. High proportions of surveyors (over 40 percent) report that respondents were unable to count up their affirmative answers to the list items without help.
>
> H4. We expect high correlations (Pearson's R <.70) across these three outcomes.

Moreover, we hypothesize that some subpopulations (e.g., differential educational attainment, age, etc.) suffer from misunderstanding this technique more than others, therefore violating aspects of the experimental design. We will interrogate these hypotheses in three ways. First, we will examine differences in which subpopulations are more likely to have revealed responses to list items, the experiment is too complicated, or help was required compared to those who did not. We will run OLS regression and binary logit with our outcome questions as the dependent variables and subpopulation indicators as our independent variables. Second, we will run similar OLS and binary logit analyses among our control group population on an indicator for whether an inconsistency occurs between direct versus list experiment counts of affirmative responses to items. This allows us to further examine whether certain subpopulations have issues with the experimental design in an indirect way. Finally, we will consider whether respondents reporting that the experiment was too complicated to answer or, alternatively, surveyors reported that respondents were unable to count up their affirmative answers without help, mediate effects of subpopulation indicators on revealing responses to specific list items. Specifically, we expect:

> H5. Less educated interviewees (primary education or below) are more likely to have difficulty responding to list experiments than higher educated interviewees.
> a) Interviewers are more likely to report that these respondents reveal their answers.
> b) These respondents are more likely to report that they find the list experiment too complicated to answer.

[11] Their study revealed 487 distinct experiments across 154 papers.

c) Interviewers are more likely to report that these respondents are unable to count up their answers without help.
d) Less educated respondents within the control group will be more likely to exhibit inconsistencies between the direct questions and list experiment items.

H6. Elderly respondents (age 60 and above) are more likely to have difficulty responding to a list experiment than younger respondents.
a) Interviewers are more likely to report that older respondents reveal their answers.
b) These respondents are more likely to report that they find the list experiment too complicated to answer.
c) Interviewers are more likely to report that these respondents are unable to count up their answers without help.
d). Elderly respondents are more likely to exhibit inconsistencies between their responses to direct questions versus the list experiment.

H7. For both of these subpopulations, we expect reporting of the list experiment as too complicated and that help was needed to count items up to mediate the primary outcome of violated anonymity.

We also suspect that confusion may adversely affect those in the treatment group in particular, who are already potentially conscientious about reporting sensitive behavior and are therefore processing additional concerns when responding to the experiment. Treatment group respondents are also dealing with one additional item than control group respondents, which could be increasing their cognitive burden leading to differential results. We note here that this our only truly experimental outcome in this study since it is the only analysis that will make use of randomization inference.

H8. Respondents within the treatment group are more likely to find responding to the list experiment difficult than those in the control group.
a) Interviewers are more likely to report that treatment group respondents reveal their answers.
b) These respondents are more likely to report that they find the list experiment too complicated to answer.
c) Interviewers are more likely to report that these respondents are unable to count up their answers without help.
d) Reporting of the list experiment as too complicated and that help was needed to count items up to mediate the primary outcome of violated anonymity.

In addition, we will explore our outcome responses by a number of other demographic and socio-economic variables, including gender, wealth, and ethnic groups. However, we do not expect significant differences between these groups since we suspect failures in list experiment implementation to arise from differential skills and faculties, not just demographics.

Finally, we expect that reporting vote-buying is not sensitive in this context. We differ from many previous list experiments on vote-buying in that we explicitly state contingency: voting in response to an offer of gifts, money or services at election time. We will test this expectation using the list experiments that interviewers and interviewees saw as implemented without problem. (That is, for which neither reported the list experiment was too complicated or involved the respondent revealing answers to the list items.) We anticipate a null effect:

H9. Interviewees will not be sensitive to revealing vote-buying. The analysis for this hypothesis will employ difference-in-means tests between the treatment and control groups of the experiment among those who did not suffer list experiment implementation failure.

**Method**

The experiment was included within a larger survey on governance and local development in Kenya, Malawi, and Zambia (N~20,000) run in April to June of 2019 (Lust et al. 2019). While the LGPI survey was only conducted in Nairobi for our Kenyan sample, it includes separate samples of both the capitals and border areas of Zambia and Malawi. Thus, we can examine the generalizability of our findings across three countries and five different regions. Treatment and control groups were randomized by programming embedded within the SurveytoGo platform, and is thus independent of the interviewer and researchers.

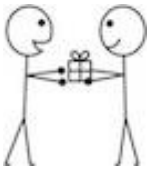
Surveyors were very well trained on the experimental protocols. We held 3-day trainings to ensure our surveyors understood the content of the surveys very well, to allow for incorporation of their comments on the instrument and its translations, and to run it a few times with their contacts first and then in a small pilot study. Researchers also spent numerous months in the field running focus groups before the survey, get to know the local teams, and ensure proper implementation of the surveys. Finally, we included very clear instructions for the list experiment, including asking enumerators to turn around while the respondent was counting up the items. We did this in order to ensure the respondent felt he/she had ample time and ability to count up the items privately. The enumerator turned around the first time the experiment is read out, in case the respondent wanted to count up the items on fingers or employ an otherwise public means to note which items he/she had engaged in. Then, the enumerator reminded the respondent that he/she should not reveal which items were engaged in and only the number, while pointing to each picture and list item, explaining each one again along with the answer options. The enumerator turned around again to allow the respondent to count up based on the pictures and select an answer on the tablet. The exact experimental prompt read as follows, with unvocalized enumerator instructions in italics and parentheses reminders only.

"Please listen to this list carefully and tell me how many of these things that you have done in the past. DO NOT TELL ME WHICH THINGS you have done in the past, ONLY THE TOTAL NUMBER of things. I will turn around to read this list to allow you to count up and answer.

(*Enumerator: Turn around and continue reading SLOWLY*)

Please count up, have you ever…

- Moved to a different village or neighborhood than this one. 

- Had water piped into your current dwelling 
- Visited with others in this village or neighborhood 
- Become more likely to vote for a candidate because they offered/gave you gifts, money, or personal favors 

(*Enumerator: Ask the respondent if they would like you to read the list again. If yes, read it again, if not, continue reading.*)

How many of these things have you done in the past? 0, 1, 2, 3, or 4? I am going to hand you the tablet now. Please press on the circle next to the number of things you have done in the past and tell me once you have done so. (*Enumerator: Hand the respondent the tablet, tell the respondent which picture corresponds with each action but remind him or her you do not want to which activity he or she has done. Turn around again, and let the respondent select an answer.*)

Beyond the traditional count outcome, we asked three additional binary questions that make up the outcomes for this study. The first probed the respondent: Do you feel that the last question was too complicated to answer properly? The second two questions were posed to the enumerator: 1) (DO NOT READ:) Respondent was able to count up the number of items without my help, and 2) (DO NOT READ:) Respondent revealed to me which items he/she had done in the past."

The experiment was designed taking into consideration of best practices in list experiment methodology. Items were selected to be straightforward, easy to recall, and with the intention of avoiding ceiling and floor effects. Ceiling effects occur when treated respondents tend to answer the maximum number of items on the list (Glynn 2013), whereas floor effects refer to the opposite occurrence: respondents in the treatment group tend to answer zero items (Blair and Imai 2012). Both effects inhibit proper analysis of list experiments. To avoid ceiling effects, we included two items that had rather low prevalence of affirmative responses in a survey we conducted in Malawi in 2016 with about 8,000 respondents (Lust et al. 2016). Moving to another village received 29 percent of positive responses among the LGPI 2016 sample, and having water piped into one's dwelling was affirmed by just 2 percent of respondents (with the Afrobarometer 2018 reporting 8, 3, and 11 percent for Kenya, Malawi, and Zambia respectively on this latter question). On the other hand, almost all respondent reported visiting others in their neighborhood or village, and only 2 percent of people denied doing this). The sensitive item asked respondents if they have ever "become more likely to vote for a candidate because they offered/gave you gifts, money, or personal favors". We also included cartoons to show to respondents in order to increase precision among lower educated

respondents (Kramon and Weghorst 2019). Finally, the order of the items were randomized to avoid recency effects (Krosnick and Alwin 1987).

Our direct questions will be coded as binaries, dropping out respondents from the study who refused to answer them or reported they did not know.

- What is the main source of drinking-water for members of your household?
    - Answer option piped water into dwelling coded as 1, other answers as 0.
- How often do you visit others in this village/neighborhood?
    - Never
    - Rarely
    - Sometimes
    - Often
    - Don't know/Refuse to answer
- How long have you lived in this village/neighborhood? Coded as all my life versus other options
    - Less than 1 year
    - More than 1 year and less than 5 years
    - More than 5 years and less than 10 years
    - More than 10 years
    - All my life
    - Don't know/Refuse to answer
- [Sensitive item] Has an offer or distribution of gifts, money, or personal favors ever made you more likely to vote for a candidate?
    - Yes
    - No
    - Don't know/Refuse to answer

**References:**


Afrobarometer Data, Malawi, Kenya, and Zambia, Round 7, 2018, available at http://www.afrobarometer.org.

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

Blair, Graeme and Kosuke Imai. 2012. "Statistical analysis of list experiment," *Political Analysis* 20(1): 47-77.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago: University of Chicago Press.

Çarkoglu, Ali and Erdem Aytaç. 2015. "Who gets targeted for vote-buying? Evidence from an augmented list experiment in Turkey," *European Political Science Review* 7, 547-566

DeMaio, Theresa J. 1984. Social Desirability and Survey. *Surveying Subjective Phenomena* 2: 257.

Glynn, Adam. 2013. "What can we learn with statistical truth serum? Design and analysis of the list experiment," *Public Opinion Quarterly* 77(1): 159-172.

Gonzales-Ocantos, Ezequiel, Chad Kiewiet de Jong, Carlos Meléndez, Javier Osorio, and David Nickerson. 2012. "Vote buying and social desirability bias: Experimental evidence from Nicaragua," *American Journal of Political Science* 56(1): 202-217.

Gonzales-Ocantos, Ezequiel, Chad Kiewiet de Jong and David Nickerson. 2014. "The Conditionality of Vote-Buying Norms: Experimental Evidence from Latin America," *American Journal of Political Science* 58(1): 197–211.

Hatchett, Shirley, and Howard Schuman. 1975. "White Respondents and Race-of-Interviewer Effects," *The Public Opinion Quarterly* 39 (4): 523–8.

Kitschelt, H., Wilkinson, S. I., et al. (2007). *Patrons, Clients and Policies: Patterns of Democratic Accountability and Political Competition*. Cambridge University Press.

Kramon, Eric and Keith Weghorst. 2019. "(Mis) measuring sensitive attitudes with the list experiment: Solutions to list experiment breakdown in Kenya." *Public Opinion Quarterly* 83(S1): 236-263.

Krosnick, Jon A., and Duane F. Alwin. 1987. "An evaluation of a cognitive theory of response-order effects in survey measurement." *Public Opinion Quarterly* 51(2): 201-219.

Krosnick, Jon A. 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys," *Journal of Cognitive Psychology* 5:213–36.

Kuklinski, J., M. Cobb, and M. Gilens. 1997. "Racial attitudes and the 'New South.'" *Journal of Politics* 59(2): 323– 49.

Lust, Ellen; Kao, Kristen; Landry, Pierre F.; Harris, Adam; Dulani, Boniface; Metheney, Erica; Nickel, Sebastian; Carlitz, Ruth; Gakii Gatua, Josephine; Jöst, Prisca; Mechkova, Valeryia; Mujenja, Maxim Fison; Tengatenga, John; Grimes, Marcia; Ahsan Jansson, Cecilia; Alfonso, Witness; Nyasente, Dominique; Ben Brahim, Nesrine; Jordan, Jenna; Bauhr, Monika; Boräng, Frida; Ferree, Karen; Hartmann, Felix; and Lueders, Hans. (2019) "The Local Governance and Performance Index (LGPI) 2019: Kenya, Malawi, Zambia." The Program on Governance and Local Development, University of Gothenburg: 2019. www.gld.gu.se

Miller, Judith Droitcour. 1984. "A New Survey Technique for Studying Deviant Behavior." PhD diss. George Washington University.

Nichter, Simeon. 2014. "Conceptualizing Vote Buying." *Electoral Studies* 35: 315–327.

Schaffer, Frederic Charles. 2007. *Elections for Sale: The Causes and Consequences of Vote Buying*. Boulder, CO: Lynne Rienner Publishers