Practical Considerations with Using Mobile Phone Survey Incentives: Experiences in Ghana and Tanzania

Robert Morello and Benjamin Leo

Abstract

As mobile phone surveys are gaining popularity among researchers and practitioners in international development, one primary challenge is improving survey response and completion rates. A common solution is to provide monetary compensation to respondents. This paper reports on our experience with using incentives with a mobile phone survey conducted in Ghana and Tanzania in June 2015.

We find that extrinsic incentives – transfers of airtime – improve survey completion rates by roughly 6 to 8 percentage points. We also find that an SMS notification, aimed at increasing intrinsic motivation to complete the survey, has a comparable positive effect. Considering three levels of compensation (airtime transfers), we find that the supply curve of completed surveys appears to be fairly inelastic. In other words, small amounts of incentives have a similar effect as larger incentives. We also find little evidence of the extrinsic incentives skewing the demographics of the completed samples. However, we do find some evidence that the SMS treatment may have skewed the sample toward wealthier and more educated respondents. Finally, we found that the cost per completed response is not substantially less for every incentive or combination of incentives, compared to the group that did not receive any incentives.

Considering these results, we would recommend the consideration of modest monetary compensation as a way of increasing response and completion rates. Nonetheless, a complete lack of extrinsic incentives may remain the most cost effective way to gather completed responses without any associated risk of skewing the survey sample. However, if higher completion rates are desired, a small monetary compensation would be most cost effective with lower risks of biasing the completed sample

Keywords: mobile phones, Ghana, Tanzania, survey design

Center for Global Development
www.cgdev.org

Working Paper 431
July 2016

Practical Considerations with Using Mobile Phone Survey Incentives: Experiences in Ghana and Tanzania

Robert Morello

Benjamin Leo
Center for Global Development

This work was co-funded and co-implemented by the World Bank's Governance Partnership Facility as part of activities of the Digital Engagement Evaluation Team (DEET). In particular, the authors thank Stephen Davenport, Tiago Peixoto, and Jonathan Mellon for their collaboration on this project.

The Center for Global Development is grateful for contributions from the William and Flora Hewlett Foundation, the Rockefeller Foundation, and the Nathan Cummings Foundation in support of this work.

Robert Morello and Benjamin Leo. 2016. "Practical Considerations with Using Mobile Phone Survey Incentives: Experiences in Ghana and Tanzania." CGD Working Paper 431. Washington, DC: Center for Global Development.
http://www.cgdev.org/publication/practical-considerations-using-mobile-phone-survey-incentives-ghana-tanzania

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

Contents

Introduction

Mobile phone surveys are gaining popularity among researchers and practitioners in international development. They offer a rapid, inexpensive way to gather citizen views and perceptions. One key concern has been that mobile phone surveys tend to have lower completion rates than those of household surveys, which have long been the industry standard. Face-to-face enumerator surveys tend to have near 100 percent completion rates because of repeat visits to respondents' households.[1] Survey methodologists often view lower completion rates as a fundamental methodological weakness, which undermines a survey's representativeness.

Although mobile phone surveys should not be viewed as substitutes for in-person enumerated, household surveys, improving survey response and completion rates is a primary challenge for researchers and practitioners. One common solution is to provide modest monetary compensation to respondents.[2] Although it is common to provide extrinsic incentives such as airtime rewards as a quick fix to deal with the non-response issue, there are significant reasons to further examine monetary and other incentives.

This paper reports on our experience with using incentives with two mobile phone surveys conducted in Ghana and Tanzania in June 2015. It is not intended as a foundational guide for how to use incentives in relation to mobile phone surveys. Instead, we hope that by reporting our practical experiences others can build upon these results in a range of country contexts.

Overall, we find that extrinsic incentives (e.g., airtime transfers) improve survey completion rates by about 6 to 8 percentage points. We also find that an SMS notification, aimed at increasing intrinsic motivation to complete the survey, has a comparable, positive effect. Considering our three levels of monetary compensation, we find that the supply curve is fairly inelastic. In other words, small amounts of incentives have a similar effect on completion rates as larger monetary incentives.

We also find little evidence of extrinsic incentives skewing the demographics of the completed samples. However, we do find evidence that the SMS treatment may have skewed the sample toward wealthier and more educated respondents. Finally, we found that

---

[1] For an example and further discussion of non-response, see Grosh, Margaret E.; Munoz, Juan. 1996. A manual for planning and implementing the living standards measurement study survey. Living standards measurement study (LSMS) working paper; no. LSM 126. Washington DC; World Bank. http://documents.worldbank.org/curated/en/1996/05/438573/manual-planning-implementing-living-standards-measurement-study-survey or see ICF International 2012. Demographic and Health Survey Sampling and Household Listing Manual. MEASURE DHS, Calverton, Maryland, U.S.A.: ICF International http://www.dhsprogram.com/publications/publication-DHSM4-DHS-Questionnaires-and-Manuals.cfm

[2] Other approaches include tactics such as conducting a baseline household survey with field enumerators and then distributing cellular phones to the sample for ongoing phone-based surveys.

incentives do not increase completion rates enough to substantially offset their costs. Therefore, not using incentives may be the most cost effective way to gather completed responses. However, if higher completion rates are desired, a small monetary compensation would be most cost effective with little risk of biasing sample demographics.

This paper is organized as follows. In section II, we outline the relevant literature, including the basis from which we pose our research questions. Section III includes our experimental study design and section IV contains a summary of our key findings. Lastly, we conclude in section V with practical recommendations for researchers and practitioners concerning the use of incentives in mobile phone surveys.

Overview of Relevant Literature

Social scientists have long studied why people engage in certain types of behaviors. Particularly, they have posited reasons why people would engage in pro-social behavior, or actions that are intended to help others at a cost to the individual that outweighs the benefit to the individual. Although we will not rehearse the years of literature and debate on this subject, we build upon this theory to consider the individual's costs and benefits of completing a mobile phone survey.

Completing a mobile survey requires a cost to the participant in terms of time and effort. Singer and Ye (2013) break out respondents' motivations to pay this survey completion cost into three categories. First, respondents may believe the research is important and their participation will benefit other people (altruistic motivation). Second, individuals may enjoy that their opinion matters (egotistic motivation). Third, respondents may be particularly interested in, or motivated by, a specific survey topic (survey characteristics). In this research, we consider all of these dynamics as a single, combined form of intrinsic motivation.

In addition to this inherent intrinsic motivation, we experimentally add two additional incentives. The first aims to enhance intrinsic motivation and the second is an extrinsic reward. Gneezy, Meier, and Rey-Biel (2011) find that extrinsic incentives have a standard price effect and indirect psychological effect. If we consider respondents as suppliers of completed surveys who are paid the price of a completed survey, these combined effects shift the price of supplying a completed survey. As with any supply function, we expect that higher compensation, either in terms of intrinsic motivation or extrinsic benefit, would lead to more completed surveys.

However, the interaction between intrinsic and extrinsic incentives tends to be more complicated and nuanced. Bénabou and Tirole (2006) explain two consequences of employing extrinsic incentives. First, their use can create doubt about the respondents' true motivation to do the pro-social behavior. In addition, they can communicate unintended

information about completion to respondents. For example, compensation may indicate that completing the survey is undesirable or difficult. Thus, an extrinsic incentive does not necessarily have a clear, additive effect with pre-existing intrinsic motivations. An extrinsic motivation potentially could undermine the intrinsic motivation to the point that there is a resulting net loss in motivation to complete the survey (e.g., a crowding-out effect). Frey Oberholzer-Gee (1997) outline this theoretical possibility in greater detail.

Singer and Ye (2013) find little evidence of this crowding-out effect in surveys. Overall, they find that incentives increase completion rates. However, they warn that there is no suggestive evidence for the optimized size of incentives. Previous studies indicate that as incentives increase, so do completion rates, although at diminishing marginal rates.

In this study, we aim to explore the magnitude of extrinsic incentives. By using three levels of extrinsic incentives in two different African country contexts, we begin to understand the slope of the supply curve for completed surveys. A very elastic curve would indicate that large increases in extrinsic incentives could substantially improve completion rates. Alternatively, an inelastic supply curve would suggest that a small extrinsic reward would have a similar effect to large ones. Put differently, larger extrinsic rewards would be largely a waste of money.

Singer and Ye (2013) warn that employing incentives may not be a panacea to address non-response bias. Since incentives are rarely, including in these surveys, targeted to populations that do not normally respond, they could actually make a survey less representative. For example, if males more positively react to an incentive than females, applying an incentive could make it more likely for males to complete a survey. Therefore, the subsequent sample would be more skewed toward males than a survey without an incentive. Thus, Singer and Ye advise caution in using incentives. In this research, we look for these unintended consequences as well.

Study Design Overview

Building upon this prior research, we used an experimental design to explore the effect of incentives in our mobile phone survey. In this section, we enumerate our three primary research questions.

Research Question 1: What is the price elasticity for completed mobile phone surveys?

Previous research (see Leo et al 2015) found that mobile phone incentives were effective in increasing completion rates but with highly contextual results. Despite the substantial (+3.4 percent points) and statistically significant increases, we found incentives were not cost-effective in a practical sense. It was more expensive to gather responses using incentives than without them because they did not increase the response rate enough to compensate for the

added expenditures. Although there may be an inherent value in higher completion rates, for example in improved representativeness, the return on investment was not high enough operationally to justify their use.

Our previous research used a single relatively arbitrary incentive level. If we would have chosen a different level, we may have found them to be cost effective or alternatively not effective at all. In order to explore this research question, we offered one of three different levels of monetary compensation to 45 percent of the total country samples. After the respondents' call connected and they selected a language, the introduction included a sentence explaining how much they would be compensated for completing the survey. Through these three levels, we begin to understand how the supply curve is shaped. In Ghana and Tanzania, 15 percent of respondents received the first level of incentive, another 15 percent received the second level, and likewise for the third level. The amounts roughly doubled between levels to plot the supply function over a spectrum of prices.

Figure 1 – Incentive Amounts[3]

	Ghana		Tanzania	
	Cedi	USD (approx.)	Shilling	USD (approx.)
Incentive Level 1	1	$0.25	500	$0.24
Incentive Level 2	2	$0.50	1000	$0.48
Incentive Level 3	4	$1.00	2000	$0.96

Research Question 2: How does intrinsic motivation compare to extrinsic motivation, in terms of increasing completion rates, cost effectiveness, and unintended consequences?

As noted previously, the literature illustrates that a combination of incentives likely motivate respondents to complete a given survey. These incentives are broadly divided into two types: intrinsic and extrinsic. We experimentally compare the extrinsic and intrinsic incentives. Appealing to respondents' sense of intrinsic motivation may substantially increase completion rates. Therefore, we want to investigate how a manipulation of intrinsic motivation may compare to extrinsic incentives.

In order to compare these two groups of incentives, half of our sample received an SMS notification the day before the survey call was placed. We designed this message to increase respondents' sense of intrinsic reward to completing the survey. It read, "You've been

[3] The referenced and different incentive levels were communicated to prospective respondents in locally denominated units of currency (e.g., Cedis or Shillings).

selected for a survey on electricity in [Ghana / Tanzania]. Please expect a call tomorrow [morning/afternoon/evening]. Add your voice to the national discussion." Following Singer and Ye (2013), we attempt to appeal in a relatively short message to the three categories of intrinsic motivation to complete a survey. First, an appeal to adding one's voice to a national discussion should remind respondents of altruistic, civic duty motivations. Second, the specific reference to one's voice should appeal to an egoistic motivation. Finally, revealing the topic on energy speaks to the third category of interest (survey characteristics). On this latter category, we acknowledge that the possibility of an opposite effect if respondents are not interested in electricity access and/or service delivery quality.[4]

We readily concede that this SMS is not the perfect operationalization of intrinsic motivation. Specifically, our SMS-based intervention was constrained by the length of the message and other operational concerns. We encourage further research in this area to better understand the effects of different categories on intrinsic motivation. Second, we acknowledge that the treatment packages a notification. Beyond our intention to increase respondents' intrinsic perception of completing the survey, other effects of prior notification are compounded into a single treatment. We are unable to determine if the observed effects are from the prior notification or from increased intrinsic motivation.[5]

Beyond comparing the completion rates between these two treatments, we also want to consider two secondary concerns. First, we identify the relative cost-effectiveness of each treatment. In a practical sense, we consider if the completion rates are increased enough to compensate for the cost of the treatment, either sending the SMS or paying the compensation. As another secondary concern, we explore if there are any unintended consequences to using these incentives. Again, Singer and Ye (2013) warn that using incentives to increase completion rates may not equate with better representativeness. If incentives have different effects across demographics groups, incentives actually could make the survey less representative. We explore this possibility for both treatments.

Research Question 3: How do intrinsic and extrinsic incentives interact?

Finally, we explore the interaction of the SMS treatment and the monetary compensation treatments. The literature presents the possibility of a crowding-out effect, where an extrinsic incentive could undermine respondents' intrinsic motivation to respond to the survey to such an extent that there is a net loss in desire to complete the survey.

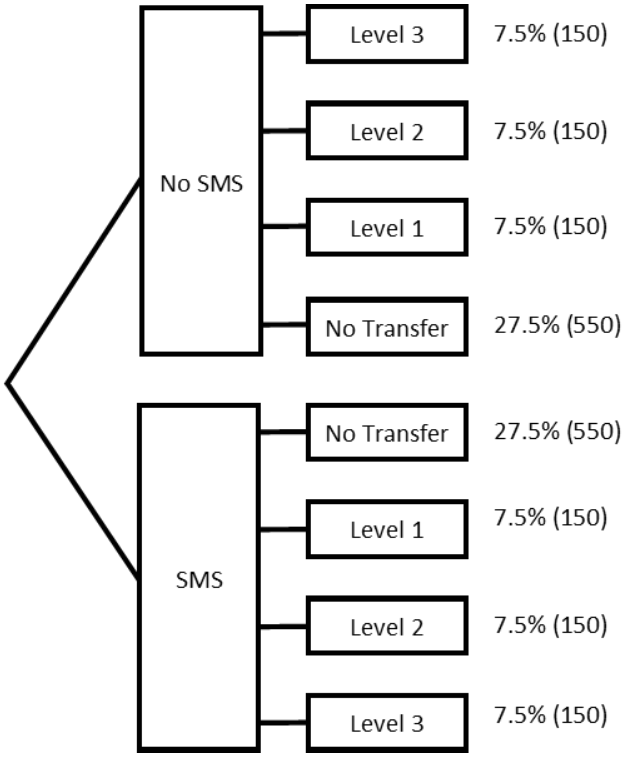
Because of the possibility of "crowding-out," we want to isolate the possibility of this effect. Evidence of crowding out would be that the response rate for the SMS group without an

[4] Importantly, these are very prominent and contentious issues in both Ghana and Tanzania; thereby, reducing the likelihood of a significant dampening impact on survey completion rates in this instance.

[5] We acknowledge the entire effect which we attribute to intrinsic motivation may be entirely due to reducing the transactional cost of completing the survey because respondents are able to plan to complete the survey.

incentive would be higher than the response rate for the group treated with an SMS and an incentive. We fully interact the treatments so that equal proportions of each compensation treatment was also treated with an SMS message. This gives us the final experimental design presented in figure 2 below.

Figure 2 – Experimental Design in Ghana and Tanzania



Results

Now that we have outlined our experimental design and our research questions, we consider the results of our surveys. First, we discuss how we analyzed the data and the evidence of properly implemented randomization. Then, we address each research question in turn.

Analysis Plan

We are most interested in the survey completion rates for each of our treatment groups. The completion rate is defined by the number of respondents in the group completing the survey divided by the total number of respondents in the respective group. We compare these means between the groups with a two-sample tests of proportions. This test makes several

assumptions about the data. First, the test assumes that our samples are independent. By virtue of our experimental design, we know that the samples are independent from each other. Second, we must have at least ten success and ten failures, which we have. Finally, our sample is much smaller than our population of interest.[6] We use a two-tailed test because we do not have a prior on whether the incentives will increase or decrease completion rates compared to the control group. Unless otherwise noted, we consider statistical significance to be a p-value less than 0.05.[7]

Evidence of Proper Randomization

In most field experiments, researchers take great care to ensure that observable attributes of participants are balanced across treatment and control groups. Blocking or stratifying procedures ensure that the participants are identical in the treatment and control groups. In this experiment, we are unable to observe any demographic attributes until after the respondents begin to answer survey questions. In order to balance these observable and non-observable attributes, we use simple randomization of all respondents into treatment and control groups.

We believe that the respondents were properly assigned into random groups. In Ghana, we find very little evidence of differences between segments. Calls were as likely to connect for numbers in the SMS treatment group as for the control. Except for one carrier, the mobile network operators were just as likely to appear in one treatment as another or in the control.[8] In Tanzania, we also find that numbers from each carrier were just as likely to appear in each treatment group. We also find that connection rates for the SMS treatment and the control are almost identical 15.4 percent for the treatment group and 15.8 percent for the control. Although this is a statistically significant difference, we conclude that this

[6] Standard rule of thumb for this assumption is that the population is normally ten or twenty times larger than the population. Both of these suggestions are met.

[7] Alternatively, we could have used a Pearson chi-squared test to compare the completion rate in a two-by-two tabulation (treatment and control by complete or not complete). When using large sample sizes with a small table, the chi-squared test is extremely sensitive to small changes in frequencies. Because of this known disadvantage of this statistic, we chose the two-sample test of proportion as our primary test for this analysis. We also ran the chi-squared test as a robustness check. Although p-value changed between the tests, we did not have any substantive changes in other interpretations of the results.

[8] The Ghanaian carrier Expresso was more likely to appear in the SMS treatment (54 percent) than in the control (46 percent). We do not believe that this unbalance comprised the overall experiment since only 0.07 percent of attempted calls on this network connected, and none of these calls actually completed the survey. Online sources report that Expresso has a very small market share, less than 1 percent. (http://www.myjoyonline.com/business/2015/April-1st/expresso-makes-its-biggest-gain-in-voice-subscriber-base.php).

difference in unlikely to affect our results because it is substantively very small. Also, it could be a spuriously correlation, a by-product of very large sample sizes.[9]

Elasticity of Completed Survey Curve

Before discussing the price elasticities of the completed surveys, we first discuss the effects of the extrinsic results. These results combine all of the 15 percent of respondents who received each incentive level. Thus, half of each level received the SMS message and half did not. As a result, any effect of the SMS message should be balanced. Also, we limit our analysis to the respondents who completed listening to the introductions and thus heard if and what quantity of incentive was offered.[10]

In aggregate, we find that an extrinsic incentive increases completion rates. In both Ghana and Tanzania, we find a substantial increase, between about 6 to 8 percentage points, that is also statistically significant. We note that the first incentive level in Tanzania caused only a modest increase of 3.4 percentage points that was not statistically significant compared to the control. We found little evidence of a statistical difference between the treatments when compared to each other.[11] In Tanzania, the second incentive level was a statistically significant increase compared to the first level.[12]

Figure 3 – Effect of Extrinsic Incentives, Survey Completion Rates

	Ghana	Tanzania
No Incentive (control)	53.1%	61.9%
Level One	59.9%**	65.8%
Level Two	61.7%**	71.8%**
Level Three	59.1%**	69.1%**

[9] In the control group, we have 214,429 phone numbers and 100,895 in the SMS treatment. In a chi-squared test, we find that the connected calls are more likely in the control with a p-value 0.005.

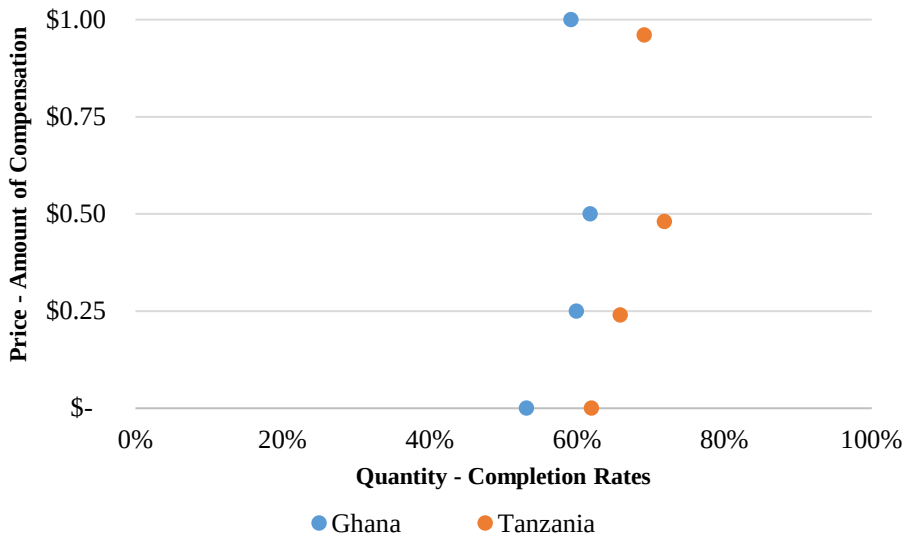
[10] Limiting our analysis to this group increases the completion rates presented in this paper because the longer respondents stayed on the survey the more likely they were to complete the survey. Because we are mainly concerned with differences in completion rates, this selection should not affect this difference among the control and treatments groups.

[11] This lack of statistical significance may be driven by smaller sample sizes, although we treated about 500 respondents in each level. In addition, we did find one statistically significant difference, noted above.

[12] The p-value was 0.04 in a two-tailed comparison of proportions test.

This initial conclusion indicates that the supply curve may be fairly inelastic.[13] Explicitly, we calculate the elasticity on different segments of the supply curve.[14] Across the entire trend, we find that both supply curves are inelastic. Specifically, we find price elasticities of 0.22 and 0.24 in Ghana and Tanzania, respectively. However, we do not find every segment to be inelastic. Between the base price of no compensation and the second incentive level, we estimate the supply curve's elasticity at 1.21 and 1.24 for Ghana and Tanzania, respectively. This portion of the curve would be considered slightly elastic.

Figure 4 – Supply Curve of Completed Surveys, by Country



Comparison of Intrinsic to Extrinsic Incentives

Next, we consider the effect of the SMS treatment, which aimed to increase respondents' perception of an intrinsic reward. We consider the entire half of respondents that received an SMS to the other half which did not. This means that just less than half of the respondents in both groups received an extrinsic incentive as well. These effects should be balanced across the groups, and the effect of the SMS can be isolated (see subsection G for a discussion of the combination effect of SMS and compensation).

---

[13] If the elasticity is calculated to equal 1, it is considered to be unitary elastic. A 1 percent increase in price corresponds to a 1 percent increase in quantity. If the elasticity is greater than 1 then, the supply curve is considered elastic. Alternatively, an elasticity of less than 1 is considered inelastic.

[14] The Elasticity of supply is defined percent change in the quantity supplied by the percent change in the price. More formally $E_s = \frac{\frac{Q_{2-Q_1}}{(\frac{Q_{2+Q_1}}{2})}}{\frac{P_{2-P_1}}{(\frac{P_{2+P_1}}{2})}} = \frac{\%\Delta Q}{\%\Delta P}$.

Before discussing the results, we note an analysis decision. We chose to conduct an intent-to-treat analysis, which is a more conservative approach. This means that we considered all numbers that should have received a SMS as part of the treated group, whether or not the respondent actually received the message. Therefore, we include the small amount of phone numbers that were unable to receive the SMS message but connected to the survey the following day, and a few of these answered survey questions. Limiting the treatment to only the numbers that actually received the SMS messages biases the completion rates up.[15] Thus, an intent-to-treat analysis gives us a more conservative perspective on the effect of SMS messages.

Now, we consider what effect the SMS treatment had on completion rates. Across specifications, we find that the SMS treatment significantly increased completion rates. For connected calls in Ghana, 8 percent that were not in the SMS treatment led to a completed survey, while 13 percent of those that were in the SMS treated did. For connected calls in Tanzania, 3 percent completed the survey, while 9 percent in the SMS treatment completed the survey. Both of these sizable increases are statistically significant.[16]

We find that the effect of the SMS message is very similar to the extrinsic incentives. In order to make this comparison, we use samples for the SMS treatment that mirror the samples for the monetary incentives. We find that SMS treatments have completion rates of 60 percent in Ghana and 69 percent in Tanzania.[17] This is compared to the control, which had a 53 percent completion rate in Ghana and 59 percent in Tanzania. The SMS treatment caused between a 7 to 10 percentage point increase in completion rates, about the same increase that we found for the extrinsic incentive treatments, which was 6 to 8 percentage points.

Cost Effectiveness of Extrinsic and Intrinsic Incentives

In a practical sense, incentives could operationally reduce implementation expenses if they increase completion rates enough to offset the associated costs. Due to this potential appeal, we consider the cost per completed response for each of our treatments. In all of our comparisons, we calculate the marginal cost of a completed response, specifically the costs

---

[15] The respondents who received the message have a higher connection rate on the day of the survey than the group who did not received the message. This indicates that receiving the message is an indication that the respondent lives in an area with better cellular coverage. The group to which the message was sent, regardless of receiving it or not, has comparable connection rates, indicating similar access to cellular infrastructure. Using an intent-to-treat analysis resolves this potentially compounding factor.

[16] Using a t-tailed t test with assumed equal variances.

[17] Respondents are not assigned into an incentive treatment until after they have selected a language. Thus, we limit the sample of the SMS treatment to respondents who answered the language selector in order to make an equitable comparison.

of the airtime, SMS notifications, and incentives. This does not include fixed costs, such as translations, audio recordings, and vendor costs. Thus, the cost per completed response is significantly lower than the total actual incurred costs. We note that here we only consider the costs of obtaining a predetermined sample size. Other qualities of the sample, such as representativeness, could outweigh this savings for many researchers.

In both Ghana and Tanzania, we find that monetary incentives substantially increase the cost per completed response. In Ghana, we spent $0.30 per completed response in the group without any compensation, while the first incentive level was $0.71 per completed response; the second level was $1.10, and the third level cost $1.99, per completed response. In Tanzania, we found that a completed response without compensation was $0.62. The first incentive level totaled $0.97 per completed response. We spent $1.27 and $2.01 per completed response for the second and third levels of compensation, respectively.

However, we seem to find a different pattern with the intrinsic incentive, where the SMS may have lowered the cost per a marginal completed survey.[18] In Ghana, the survey vendor spent $0.004 on each SMS message, and spent $0.006 on each message in Tanzania. The average cost per completed SMS response in Ghana was $0.297, while the average without an SMS was about $0.305, just $0.007 different. In Tanzania, we spent $0.64 per completed response in the control and spent $0.61 per complete response in the SMS treatment, a difference of about $0.027.[19] These differences are extremely small and may not hold with larger samples. The survey vendor also purchases bulk SMS in Euros, and thus the small improvements in cost could be lost with small changes in exchange rates.

In terms of pure cost considerations, we recommend either no incentives or small to modest monetary compensation. Using no incentives appears to be just as cost effective for gathering completed responses as using SMS messages, and less expensive than using monetary compensation. If a higher response rate were desired, then a small monetary compensation amount would be the most cost effective because of the inelastic supply of completed surveys. As noted previously, we found that the SMS message and the compensation comparably increased completion rates. However, the SMS treatment may have skewed our sample to be less representative of the national population (see subsection F below).

---

[18] We readily concede that this SMS treatment is not the perfect operationalization of intrinsic motivation. For further discussion, please see the second research question in section III.

[19] This is especially surprising in Tanzania because respondents with the SMS notification on average tend to spend longer responding to the survey. We find that Tanzanian respondents in the SMS treatment and the control each spent about 5 minutes to complete the survey. However, respondents in the SMS treatment who do not complete the survey spent 20 more seconds on the survey. This leads to additional costs without increasing the number of completed responses. In Ghana, we do not find this pattern.

Figure 5 – Average Marginal Cost of Completed Response, by Treatment and Country

	Ghana	Tanzania
No Compensation	$0.30	$0.62
Level 1	$0.71	$0.97
Level 2	$1.10	$1.27
Level 3	$1.99	$2.01
No SMS	$0.305	$0.64
SMS	$0.297	$0.61

Unintended Consequences of Incentives

Singer and Ye (2013) warn that using incentives to increase completion rates may not equate with better representativeness. If incentives have different effects across demographics groups, they could make the survey less representative. We check for this possible unintended consequence.[20]

We find some evidence that the incentives created unintended sampling consequences, especially for the SMS treatment in Ghana. We look for incentives skewing the demographics by comparing the completed samples in each of the treatment groups to the control group. If the incentives did not skew the demographics, we would expect that the demographics of the completed samples would be the same in the control and all treatment groups.

On three demographic components – location (e.g., urban/rural), age, and education – we compare the distribution of the control to the distribution of the treated groups, using a Pearson chi-squared test for equality of distribution. We find a statistical difference, with p-values less than 0.05, on location in Ghana and Tanzania. And, we find a difference in distribution on education in Ghana. However, this test does not identify which category or categories of a distribution are different or in which direction the treated group may differ from the control.

[20] In this section, we only explore whether incentives change the demographics of the resulting samples. We do not make value judgements on whether this effect is good or bad for the sample. It is possible that incentives may skew a sample of mobile phone respondents to be more like the general population (e.g. poorer). We do not further consider these questions here. For further consideration on if mobile surveys are representative, please see Leo et al. 2015. "Do Mobile Phone Surveys Work in Poor Countries?" CGD Working Paper 398. Washington, DC: Center for Global Development. http://www.cgdev.org/publication/do-mobile-phone-surveys-work-poor-countries-workingpaper-398

To further explore these dynamics, we use a two-group proportion-comparison test to analyze if the mean of the control sample is statistically different from the treated samples. In this analysis, we compare the means of eighteen demographics in Ghana and seventeen demographics in Tanzania.[21] When we use this many tests, there is a risk of committing a type I error, in other words identifying a statistically significant difference when one actually does not exist. In this case, we have about a 60 percent chance of committing such an error. In order to correct for this possibility, we use a Bonferroni correction as a robustness check.[22] This correction adjusts the standard for statistical significance downward to account for testing multiple indicators of a single hypothesis.[23]

It appears that the SMS treatment in both Ghana and Tanzania may skew the samples toward wealthier and more educated respondents. The results for education mirror what we found in the Pearson chi-squared test. In Ghana, the control sample includes 16 percent of respondents with more than a secondary education, while the SMS sample has 21 percent of respondents with this respective education level. Also, we found a 6 percentage point increase in the percent of respondents with finished walls in the SMS treatment group compared to the control sample (e.g., one of two asset ownership questions that were included as proxies for overall wealth). However, we did not find a significant difference between the samples in terms of ownership of a television. In Tanzania, we identify a similar, yet weaker, pattern. We found a 5 percentage point increase in the mean of respondents with finished walls and a 5 percentage point increase in terms of possessing a bank account in the SMS treatment compared to the control group. We also found a three percentage point increase for respondents with more than a secondary education.[24] In Ghana, all of these differences are robust to the Bonferroni correction. This is not the case in Tanzania.

---

[21] Location (urban, rural), ownership of two assets, gender, five age categories, and six levels of education. In Ghana, we used four location demographics: Accra, Kumasi, rural, and other urban. In Tanzania, we used three location demographics: Dar es Salaam, rural, and other urban area. This accounts for the difference in the number of tests between the two countries.

[22] We note that using the Bonferroni correction makes it easier to obtain our desired result which is not finding a change in demographics, e.g. not rejecting the null hypothesis of no difference. For this reason, we present both whether the result was statistically significant at the 0.05 level and if it was robust to the Bonferroni correction. See Appendix I for more detailed results.

[23] The Bonferroni correction divides the critical value α by the number of tests. In Ghana we have 18 tests, so we divide the required p-value 0.05 by 18 to have a corrected p-value of 0.0028. In Tanzania, we have 17 tests, so our corrected p-value is 0.0029. Small corrections to the proportional decrease can be made if the outcome variables are correlated. For example, living in an urban area could be positively correlated with the measures of wealth. Accounting for these correlations, adjusts the corrected Bonferroni level slightly upward than what is reported here. However, we do not find a substantive difference on if the test statistics are robust to the correction.

[24] We find a mixed story on location. In Ghana the control sample has 5 percent points more respondents from Kumasi, an urban area. In Tanzania, we find the control sample has about a 6 percentage point increase over the SMS treatment in the percent of rural respondents. The Tanzania result is not robust to the Bonferroni correction, while the result from Ghana is. We suspect that a local effect with Kumasi is occurring in Ghana, and we do not suspect a strong pattern of the SMS treatment affecting locality.

We do not believe that the extrinsic incentives skewed the demographics of the ultimate sample. Although we found some evidence of differences between the control and treatment groups, these differences appear inconsistent. For example, we find about 49 percent of respondents own a bank account in the control. At the 500 shilling and 1500 shilling levels, we find that about 42 percent of respondents have bank accounts. Thus, these populations seem slightly less wealthy than the control, and perhaps we could conclude that offering monetary incentives encourages poorer respondents to complete the survey. However, we hesitate to make this conclusion because we did not find differences on the other included wealth measure (possession of finished walls). Also, none of the differences are robust to the Bonferroni correction. Therefore, we suspect that these differences may be the products of spurious correlations.

Combination of Extrinsic and Intrinsic Incentives

To understand the effects of combining the intrinsic and extrinsic incentives, we fully interacted our treatments. Thus, about one in five respondents received both the SMS and were offered some type of monetary compensation for completing the survey. We now consider the completion rates of these combination treatments.

We find that the combination of intrinsic and extrinsic incentives had a tremendous positive impact on completion rates. Specifically, completion rates were between 12 and 20 percentage points higher compared to the respondents who did not receive an SMS or any monetary incentives. Using a two-tailed comparison of proportions test, we find that these increases are all statistically significant.

We fail to find evidence of a crowding out of intrinsic motivation by extrinsic incentives. If the airtime compensation crowded out intrinsic motivation, we would expect that the combination effects would have lower completion rates than the SMS-only treatment. However, we find the opposite pattern, whereby each combination treatment actually has a higher completion rate (ranging from 4 to 11 percentage points) than the SMS treatment alone. However, none of these increases are statistically significant, with one exception noted below. In both Ghana and Tanzania, we find 8 and 11 percentage point increases respectively for the second incentive level, which also are statistically significant.[25] However, since the trend is not stable across all of the combinations, we do not conclude that adding monetary compensation to the SMS treatment necessarily increases completion rates.

[25] We also find that the 9 percentage point increase in Tanzania for the third level of incentive is statistically significant.

Figure 6 – Effects of SMS and Compensation treatments, by Country

	Comparison	Completion Rate	Percentage Point Increase from Comparison		
			Level 1 & SMS	*Level 2 & SMS*	*Level 3 & SMS*
Ghana	Base *(no SMS and no compensation)*	49%	+14**	+16**	+12**
	SMS Only *(no compensation)*	57%	+ 7	+ 8**	+ 4
Tanzania	Base *(no SMS and no compensation)*	57%	+13**	+20**	+18**
	SMS Only *(no compensation)*	65%	+ 4	+11**	+ 9**

** P-value < 0.05 in a two-tailed two-group proportion-comparison test

Alternatively, we could consider the combination of incentives by analyzing the effect of adding an SMS message to a level of compensation. Using a two-tailed two-group proportion-comparison test, we compare the completion rates by different incentive levels to the group of respondents who received the same level of incentive as well as a SMS notification. Again, we find evidence that the combination of incentives is more powerful than the incentives by themselves. The combination of incentives contributes to completion rates that are between 4 and 13 percentage points higher than the incentive levels by themselves.

In Ghana, these increases tend to be more moderate, between 4 and 7 percentage points, and are not statistically significant at the 0.05 level.[26] In Tanzania, the increases ranged between 11 and 13 percentage points and are statistically significant. We suspect that the combination of incentives may be more powerful in Tanzania for some contextually specific reasons.

[26] The p-value is 0.12 for the first level of incentive and SMS compared to the first level without SMS. It is 0.14 for the second level with SMS to the second level without SMS, and likewise the third level is 0.35.

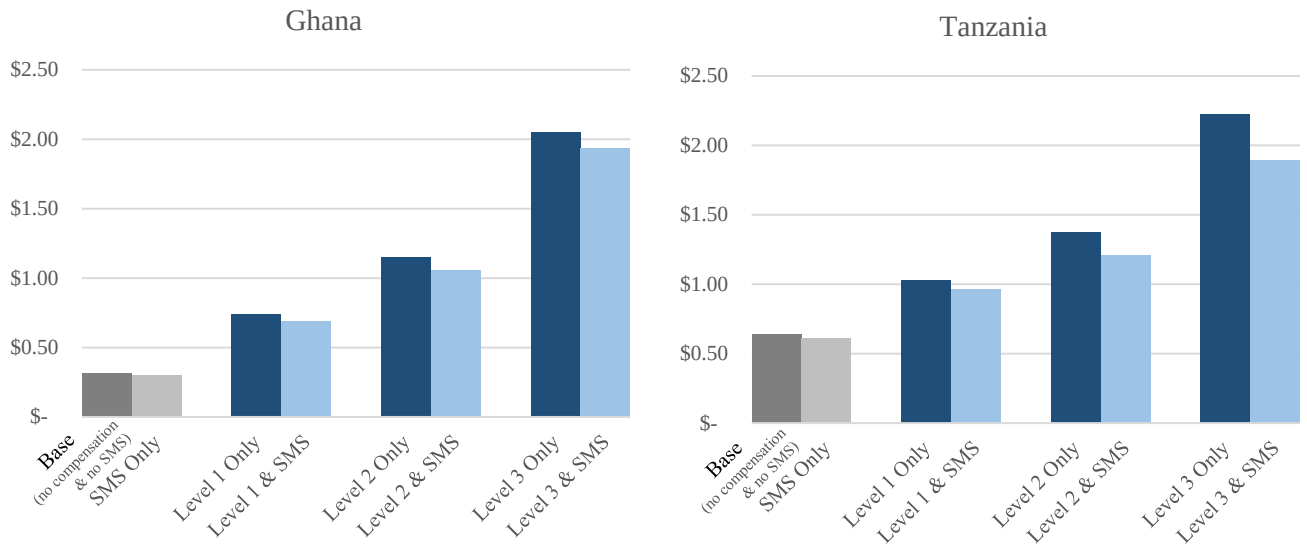
Figure 7 – Additive Effects of SMS on Levels of Compensation, by Country

	Ghana			Tanzania		
	Completion Rate (%)	Percentage Point Increase (%)	P-value < 0.05	Completion Rate (%)	Percentage Point Increase (%)	P-value < 0.05
Base *(no SMS and no compensation)*	49			57		
SMS Only *(no compensation)*	57	+8	**	65	+9	**
Level 1 Only *(no SMS)*	57			59		
Level 1 & SMS	63	+7		70	+11	**
Level 2 Only *(no SMS)*	58			65		
Level 2 & SMS	65	+7		77	+12	**
Level 3 Only *(no SMS)*	57			61		
Level 3 & SMS	61	+4		75	+13	**

** P-value < 0.05 in a two-tailed two-group proportion-comparison test

We find that the combination incentives are not cost effective, which reiterates the earlier findings related to individual incentives. Although the combination incentives seem to increase completion rates, the cost per completed response across treatments remains more expensive than comparison groups. For example in Tanzania, the third incentive level and SMS treatment increased completion rates by 18 percentage points over the group that did not receive a message or any compensation. However, this combination treatment entailed costs of almost three times as much per completed response ($1.89 versus $0.64).

Figure 8 – Cost of Combination Incentives, by Country



Conclusion

As mobile phone surveys are gaining popularity among researchers and practitioners in international development, one primary challenge is improving survey response and completion rates. A common solution is to provide monetary compensation to respondents. This paper reports on our experience with using incentives with a mobile phone survey conducted in Ghana and Tanzania in June 2015.

We find that extrinsic incentives – transfers of airtime – improve survey completion rates by roughly 6 to 8 percentage points. We also find that an SMS notification, aimed at increasing intrinsic motivation to complete the survey, has a comparable positive effect. Considering three levels of compensation (airtime transfers), we find that supply curve appears to be fairly inelastic. In other words, small amounts of incentives have a similar effect as larger incentives. We also find little evidence of the extrinsic incentives skewing the demographics of the completed samples. However, we do find some evidence that the SMS treatment may have skewed the sample toward wealthier and more educated respondents. Finally, we found that the cost per completed response is not substantially less for every incentive or combination of incentives, compared to the group that did not receive any incentives.

Considering these results from our experience, we would recommend the consideration of modest monetary compensation as a way of increasing response and completion rates. Nonetheless, a complete lack of extrinsic incentives may remain the most cost effective way to gather completed responses without any associated risk of skewing the survey sample. However, if higher completion rates were desired, a small monetary compensation would be most cost effective with lower risks of biasing the completed sample.

References

Bénabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behavior." American Economic Review, 96(5): 1652-1678.

Frey, Bruno S. and Oberholzer-Gee, Felix. "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding- Out." The American Economic Review, Vol. 87, No. 4 (Sep., 1997), pp. 746-755.

Gneezy, Uri; Meier, Stephan; Rey-Biel, Pedro. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." Journal of Economic Perspectives. Vol. 25, No. 4, Fall pp. 191-210.

Grosh, Margaret E.; Munoz, Juan. 1996. A manual for planning and implementing the living standards measurement study survey. Living standards measurement study (LSMS) working paper; no. LSM 126. Washington DC; World Bank. http://documents.worldbank.org/curated/en/1996/05/438573/manual-planning-implementing-living-standards-measurement-study-survey

ICF International 2012. Demographic and Health Survey Sampling and Household Listing Manual. MEASURE DHS, Calverton, Maryland, U.S.A.: ICF International http://www.dhsprogram.com/publications/publication-DHSM4-DHS-Questionnaires-and-Manuals.cfm

Leo, Ben; Morello, Robert; Mellon, Jonathan; Peixoto, Tiago and Davenport, Stephen. 2015. "Do Mobile Phone Surveys Work in Poor Countries?" CGD Working Paper 398. Washington, DC: Center for Global Development. http://www.cgdev.org/publication/do-mobile-phone-surveys-work-poor-countries-workingpaper-398

Singer, Eleanor, and C. Ye. 2013. "The Use and Effects of Incentives in Surveys." Annals of the American Academy of Political and Social Science, 645(1): 112-141.

Appendix

Demographics of Completed Samples, SMS Treatment and Control

Ghana					
Category	**Demographic**	**Control (%)**	**SMS (%)**	**P-value < 0.05**	**Robust to Bonferroni Correction**
Location	Accra	27	29		
	Kumasi	17	12	**	Yes
	Other Urban	19	21		
	Rural	37	38		
Wealth	Finished Walls	72	78	**	Yes
	Owns Television	76	78		
Gender	Male	64	64		
	Female	36	36		
Age	Younger than 15	10	9		
	15 - 24	50	53		
	25 - 34	28	25		
	35 - 55	9	8		
	More than 55	4	5		
Education	No Formal Education	11	9		
	Some Primary Education	20	17	**	
	Primary Complete	21	21		
	Some Secondary Education	16	15		
	Secondary Complete	17	17		
	More than Secondary	16	21	**	Yes
Tanzania					
Location	Dar es Salaam	22	25		
	Other Urban	28	31		
	Rural	50	44	**	No
Wealth	Finished Walls	66	70	**	No
	Owns Bank Account	43	48	**	No
Gender	Male	74	73		
	Female	26	27		
Age	Younger than 15	2	3		
	15 - 24	52	53		
	25 - 34	27	28		
	35 - 55	14	13		
	More than 55	4	3		
Education	No Formal Education	6	5		
	Some Primary Education	29	28		
	Primary Complete	19	20		
	Some Secondary Education	14	14		
	Secondary Complete	22	20		
	More than Secondary	10	14	**	No

Demographics of Completed Samples, By Extrinsic Incentive Level

Category	Demographic	No Incentive (%)	Level 1 Incentive	P-value < 0.05	Robust to Bonferroni Correction	Level 2	< 0.05	Robust to B.C.	Level 3	< 0.05	Robust to B.C.
Ghana											
Location	Accra	**28**	26			31			34		
	Kumasi	**14**	15			16			15		
	Other Urban	**21**	21			17			20		
	Rural	**37**	38			36			32		
Wealth	Finished Walls	**75**	76			77			78		
	Owns Television	**76**	77			77			82		
Gender	Male	**64**	59			63			66		
	Female	**36**	41			37			34		
Age	Younger than 15	**8**	9			7			7	**	No
	15 - 24	**52**	53			52			45	**	No
	25 - 34	**26**	26			27			32		
	35 - 55	**9**	9			7			7		
	More than 55	**5**	3			5			5		
Education	No Formal Education	**9**	11			8			11		
	Some Primary Education	**18**	18			21			18		
	Primary Complete	**22**	22			16	**	No	19		
	Some Secondary Education	**15**	14			17			15		
	Secondary Complete	**18**	16			15			16		
	More than Secondary	**18**	18			23			21		
Tanzania											
Location	Dar es Salaam	**24**	26			24			22		
	Other Urban	**29**	29			32			32		
	Rural	**47**	45			44			46		
Wealth	Finished Walls	**69**	64			71			70		
	Owns Bank Account	**49**	42	**	No	45			42	**	No
Gender	Male	**74**	72			73			75		
	Female	**26**	28			27			25		
Age	Younger than 15	**2**	3			2			2		
	15 - 24	**53**	49			57			52		
	25 - 34	**27**	29			32			32		
	35 - 55	**14**	14			11			13		
	More than 55	**3**	4			4			2		
Education	No Formal Education	**5**	6			5			5		
	Some Primary Education	**27**	33	**	No	26			30		
	Primary Complete	**19**	17			22			19		
	Some Secondary Education	**14**	13			16			12		
	Secondary Complete	**21**	19			20			22		
	More than Secondary	**14**	12			11			11		