The effects of monetary and non-monetary incentives on respondent attrition in longitudinal surveys

by

Ngonidzashe Nicholas Choga

BSc Computer Science, University of Cape Town, 2014

BSc (Hon) Computer Science, University of Cape Town, 2015

Dissertation presented for the degree of Master of Science

in the Department of Computer Science

University of Cape Town

January 2019

Supervised by:

Dr. Brian DeRenzi

Dr. Todd Rosenstock

I know the meaning of plagiarism and declare that all of the work in the dissertation, save for that which is properly acknowledged, is my own.

Signed by candidate

07/02/201

Mr. Ngonidzashe Nicholas Choga

Date

Abstract

Longitudinal studies are essential for governments and organizations as they help in making decisions that are based on factual data. Longitudinal studies collect data repeatedly from a set of participants over a period of time, enabling the tracking and studying of entity behaviour at individual, organizational, and national levels. One major challenge facing longitudinal data collection is the attrition of subjects during the course of the study, which is the continuous loss of participants during a longitudinal survey due to verbal drop-outs and non-response. Attrition can render datasets useless due to incomplete entries, making it one of the most significant weaknesses of longitudinal surveys. In order to explore the effects of incentives on attrition, this research project studies the effects of monetary and non-monetary incentives on explicit (subject says s/he does not want to be part of the study) and implicit (non-response) attrition. In particular, this study uses telephonically delivered feedback, derived from participant responses, as non-monetary incentives. To measure the effects of incentives on attrition, the study gave four treatments groups —50 participants each—mobile credit, verbally delivered feedback, machine delivered feedback and no incentive. After monitoring their attrition, over a 12-week period that involved bi-weekly surveys, a generalised linear model and Cochran's q-test were used to find that monetary incentives remain the strongest in under-served community settings. It was not only found that monetary incentive treatments completed the most surveys most weeks, but also had the least explicit attrition. Surprisingly it was also found that machine delivered feedback performed similarly to mobile credit when the cost, social impact and participant behaviour in terms of their survey completion and attrition is assessed.

Dedication

I dedicate this dissertation to my mother.

Without You
Mom, without you, there would be no me.
Your love, your attention, your guidance,
have made me who I am.
Without you, I would be
lost, wandering aimlessly,
without direction or purpose.
You showed me the way
to serve, to accomplish, to persevere.
Without you, there would be an empty space
I could never fill, no matter how I tried.
Instead, because of you,
I have joy, contentment, satisfaction and peace.
Thank you, mom.
I have always loved you and I always will.

By
Joanna Fuchs

Acknowledgements

This dissertation has given me a number of highs and lows as is expected. Without all the people that helped me through this journey, I would not have been able to get through the process.

Firstly, I would like to thank my energetic and witty supervisor, Dr. Brian DeRenzi. In the lowest moments, he was always there to calm me down and keep me going. His dedication to quality work has made me better in all aspects of life. I'd like to thank the principal investigator on the SCAN project, Dr. Todd Rosenstock. He was always full of knowledge whenever his guidance and experience were required. I'd like to thank Anthony Poon as his presence on early field trips helped keep me calm in the field.

I'd like to thank The African Climate & Development Initiative. They were quick to say when we required a partner for our project. In particular, I'd like to thank Nadine Methner, Likho Sikutshwa and Penny Price. I'd also like to thank all my friends from the FLOW program, Ian, Loritha, Ilichia, Emichia, Thamryn and Camy.

Lastly, I'd like to acknowledge everyone in the Centre of ICT4D at the University of Cape Town. Over the years, each and every one of them has helped me in one way or another. To everyone else who helped on this project, I say thank you.

Publications

Early work can be found in the following publication:

1. (Extended Abstract) Ngonidzashe Nicholas Choga, Brian DeRenzi, Todd Rosenstock: The effects of monetary and non-monetary incentives on respondents of longitudinal surveys, HCI Across Borders Symposium, ACM conference on Human Factors in Computing Systems (CHI), Denver, Colorado, US, CHI 2017.

Table of Contents

Table of figures

Table of tables

1 Introduction

The fundamental act of tracking change over time is the essential building block for a data-driven society. Without understanding how mortality rates, incomes, employment rates and global warming affect society, it would not be possible to accurately gauge the effectiveness of large-scale interventions. In order to achieve this, longitudinal data is most often used. A longitudinal survey is a tool that collects data from the same sample of participants on multiple occasions over a period of time. By contrast, a cross-sectional survey gathers data at only one particular point in time [49]. At an individual level, it is impossible to study dynamic phenomena, such as how a person's opinion changes over time, without using longitudinal data gathering. At an organisational level, it is very difficult to make data-driven decisions about resource allocation and service delivery without the use of longitudinal data. At a national level, the census is one of the most important government processes in the history of the New World, as it not only keeps track of changes in population but also tracks human migration patterns [54]. Longitudinal surveys are essential for governments and organisations as they help in making decisions that are based on factual data.

This study prioritizes longitudinal surveys because monitoring is essential to ensure accurate data collection in so many situations today. Well-collected data can help monitor successes and identify areas of weakness, as well as help assess the effectiveness of current practices at an individual, organizational and national level [77]. Data that is not well-collected can result in challenging data analysis and in extreme cases, useless data [6]. Menard asserts that longitudinal surveys are "ultimately indispensable" in the monitoring of the change process [54]. This project is a first step in monitoring an underserved community's current conditions with the study's feedback serving as a catalyst for change.

1.1 Challenges of collecting longitudinal data

Although critical to inform our understanding, in practice, longitudinal studies face many challenges. Longitudinal surveys are typically expensive and are often time intensive. Longitudinal

surveys also suffer from attrition [49]. Attrition is both organizational and methodological [35]. It is an organisational challenge due to the effort required to participate in longitudinal surveys. It is methodological due to the effect attrition has on the sample of participants. In other words, attrition has the potential of skewing results. Despite the negative aspects of longitudinal surveys, Menard claims, when faced with a choice between cross-sectional and longitudinal research, "the choice should be between doing the research properly or not doing it at all" [54].

Of the many challenges that collecting longitudinal data presents, including their time-consuming nature, high costs and demand for large sample groups, Thomas et al. refers to attrition in particular as the "Achilles heel" of longitudinal surveys [71]. Thus, tools that mediate this weakness are necessary. The most common tool for repeated-measures analysis ANOVA (analysis of variances) is unable to handle missing data and therefore incomplete data would have to be thrown out completely [35]. Due to the nature of longitudinal data, attrition predominantly leads to a decrease in the quality of collected data. As a result, researchers spend time and effort attempting to minimize attrition and its effects [49, 43].

Attrition in underserved regions, in most cases, depends on the needs of the community. Hill highlights this reality when he concludes that the most appropriate method of dealing with attrition depends on the aims of the study in relation to the needs of the community [43]. Generally, in an attempt to reduce attrition in underserved regions, researchers primarily provide respondents with incentives in the form of monetary compensation or gifts. These have been shown to have positive effects on the rate of attrition, however, they also have a negative effect in the form of the expectation that is built in respondents [49].

1.2 The study

In order to study the challenges and incentives of longitudinal studies, we designed a one-factor, between-subjects' study with four levels, consisting of four treatment groups, which was run over a twelve-week period. A fortnightly survey was used to collect data, with incentives delivered fortnightly on weeks after a survey. The treatment groups consisted of four groups of participants located in physically separate but socioeconomically similar locations. Each of these treatments

received a different incentive for survey participation, namely: mobile credit, a voice call delivering feedback, or an automated recording which also delivered feedback. The control group received no incentive. The treatment groups consisted of between 48 and 52 participants each. The recruitment inclusion criteria in the four locations included households that received social grants but also had additional sources of income such as seasonal work or a part-time job.

1.3 Research Questions

This study was designed to explore and answer 3 major questions:

1. How do different incentives affect attrition?
2. How do different incentives relate to a participant's habits during a longitudinal survey?
3. How do different incentives affect the cost of a project?

The first question was designed to answer the most pertinent point about attrition research because an incentive in a longitudinal survey is used in order to reduce the attrition rate of a project and as described in section 1.2, reducing attrition is key to having usable data [49]. In order to answer this question, attrition is broken down into two parts: hard attrition, which is when people drop out of the study completely, and soft attrition, which includes people that verbally ask to skip a survey week completely.

The second question is made up of a set of activities that encapsulate all possible participant behaviours. Six different participant types were defined, namely, active participants, active drop-outs, wafflers, low participation drop outs, drop outs, and non-participants.

The last question responds to a very important issue for researchers. How much does it all cost? It was found that cost was used as a reason for not providing feedback for most research projects, as established by Cox et al [20]. This begs further investigation into which is more expensive, feedback or traditional incentives such as monetary gifts? Understanding this will enable future researchers to make educated decisions when attempting to reduce rates of attrition in a manner

that will not financially affect a project's completion. Furthermore, understanding this will help researchers identify whether the cost of specific incentives is worth the reduction in attrition.

1.4 Results

The study found that compared to the control, both monetary incentives and feedback had positive effects on attrition. It found that attrition was highest in the control group, followed by the voice group, the automated group and the airtime group in that order. It was surprising that the automated feedback had a greater effect than the voice feedback due to a lack of interaction with the automated feedback. Prior research has shown that human interaction has a positive effect on the participant receiving the call [82].

The study also found that the monetary incentive and automated feedback resulted in high numbers of active participants, whereas voice feedback resulted in a high number of members that were not active but also did not explicitly drop out. As expected, the control group had a high number of drop-outs due to the lack of incentive.

Monetary incentives proved to be one of the least expensive in terms of both finances and time. This demonstrates why researchers favour them over other incentives [46, 18, 5]. The study did, however, find that automated feedback had similar effects on attrition, lower financial cost and higher time cost in comparison to monetary incentives.

1.5 Partners

The research conducted required partners in order to navigate communities with a deeper understanding of their culture. The researchers, therefore, partnered with three organisations.

1.5.1 Surveillance of Climate-smart Agriculture for Nutrition (SCAN)

SCAN is a collaboration between the University of Cape Town, specifically, the Centre for Information and Communication Technology for Development (ICT4D) and World Agro-Forestry Centre (ICRAF). The Centre for ICT4D is a part of the Computer Science department. SCAN is a project that was set up to sustainably increase production and improve the resilience of food systems under climate change [61]. It also serves as the umbrella project under which this research project is located.

SCAN provided financial and advisory assistance for this project, which was initially planned for remote communities in Kenya, however due to reasons beyond our control, the location was moved to South Africa. SCAN therefore proposed working with an initiative called African Climate and Development Initiative as they were conducting projects in Western Cape and thus a collaboration was established.

1.5.2 The African Climate and Development Initiative (ACDI)

ACDI is an initiative based at the University of Cape Town (UCT) that serves as an active response to climate change and development challenges. This project aims to support collaborative research and train communities on climate change and development. ACDI was therefore an appropriate partner for this project due to their focus on climate change, sustainability and local community development.

This partnership came at an appropriate time as ACDI was planning on running pilot surveys on food, water, energy and sustainability in the Western Cape in preparation of a larger project set to start in 2018. A partnership was therefore formed in which surveys could be run for them while collecting data on the attrition of various communities. ACDI also offered to help with the construction of feedback that would be given to participants after the surveys.

1.5.3 Fostering Local Wellbeing (FLOW)

In order to run pilot surveys in the Western Cape, ACDI was planning on working with an organization it started called Fostering Local Wellbeing (FLOW). FLOW is as local organization

located in the Bergrivier Municipality in the Western Cape. The project recruits local young adults to become FLOW Ambassadors and helps them build skills, such as video editing, data collection and general farming practises. The training allowing them to help their communities regardless of the growing challenges of climate change, resource depletion and inequality. In short, FLOW provides young adults with a one-year learner-ship program.

FLOW was an appropriate partner for the researchers because of their knowledge of the Bergrivier Municipality where surveys were to be run. They were able to identify the location of the study's desired sample group as well as guide the researchers through the best approach when attempting to recruit participants. The partnership also proved fruitful as the FLOW ambassadors were utilized as enumerators on this project. In it, enumerators recruited participants as well as conducted surveys. FLOW allowed them to assist on this project because it equipped them with additional skills that they did not necessarily gain during their learnership. We compensated the FLOW Ambassadors monetarily for their participation in the project on a weekly basis.

1.6 Context

This research project was run in the Western Cape, South Africa. In order to satisfy our methodological needs, four towns were selected in which surveys would be conducted. The research process taken is detailed in the following chapters of this dissertation. The four locations were found with the help of Fostering Local Wellbeing (FLOW), as their coordinator is vastly knowledgeable about the towns in the Bergrivier Municipality. This particular municipality was chosen in order to capitalise on the knowledge of the coordinator as well as to build on the close relationship the researchers had with FLOW and The African Climate and Development Initiative (ACDI). The similarities in all the communities in the municipality also made it ideal as four locations could be chosen with similar demographics. Lastly, the survey was meant to be conducted with marginalized communities that receive social grants but also have additional sources of income such as seasonal work or a part-time job, which is the situation with large numbers of community members in the municipality. Within the four towns, communities were pin-pointed in which surveys would be run. With the assistance of ACDI, there was a scope in which participants could be identified.

2 Background and literature review

To inform our exploration of incentives in longitudinal studies, we consulted literature to assess their advantages, disadvantages and survey design considerations. The literature will also help examine attrition and ways in which it can be minimised. This is done in order to present past work and findings in the area of longitudinal surveys, attrition and participant incentives. The information presented will be essential as it will help us understand the decisions that were made in the design of the study.

2.1 Defining longitudinal surveys

Longitudinal surveys have proven to be a powerful resource for research in economics and the social sciences [71]. Organizations—corporate, academic and social—use longitudinal surveys to monitor changes in events over a period of time. “A longitudinal survey is one that collects data from the same sample elements on multiple occasions over time”. The longitudinal surveys conducted for this project are more specifically called household panel surveys. Lynn classifies them as surveys that are multi-topic and can be used for the general purpose of collecting behavioural, attitudinal and circumstantial data on a range of socio-economic issues [49]. Longitudinal surveys are the only way to investigate how a person or an organization changes its standpoint about a specific matter over time. Furthermore, they allow us to investigate reasons for the change, or the extent of the change [35].

Longitudinal surveys on individuals and households are particularly useful when monitoring how living standards change over time [27]. This makes longitudinal surveys particularly attractive when compared to cross-sectional surveys which captures people’s thoughts in one particular instance.

2.1.1 Advantages of longitudinal surveys

Longitudinal surveys have many advantages such as their ability to monitor change, observe trends and collect data of relatively highrt accuracy. The first advantage of longitudinal surveys is their

ability to allow for the analysis of gross change as well as the analysis of unit-level change [58, 49]. Repeated cross-sectional surveys on the other hand can only estimate net change. In the case of our research, repeated cross-sectional surveys would only be able to help us identify the proportion of the community that regularly goes without food versus longitudinal surveys that would help us identify the factors affecting food shortages and how people are affected over time.

The second advantage of longitudinal surveys involves their ability to provide aggregate measures. Aggregate measures are required in order to analyse gross and unit change. Deriving measurements from different time points can only be achieved using longitudinal surveys. Aggregate measures also help with measuring stability or instability. When organizations are going through structural changes, policy makers can make informed decisions based on aggregate measures [49].

The third advantage of longitudinal surveys involves how they capture temporal characteristics of events or circumstances. Temporal events and circumstances are best captured using longitudinal surveys and are particularly important to development projects like ours. Development projects often deal with marginalization, unemployment and food insecurity which are often time sensitive [51]. A relevant example is seasonal farm workers in the Western Cape winelands who are unemployed half of the year and therefore their vulnerability fluctuates throughout the year.

Longitudinal surveys are also advantageous as they help us identifying causality. A key attribute in decision making in organizations is identifying causality. This is achieved by the monitoring of trends and identifying change. The mere identification of association is often not enough for policy change [44].

The fifth advantage of longitudinal surveys is their ability to increase data accuracy when compared to other survey formats. This is as a result of them not requiring participants to retrospectively recall details. Longitudinal surveys limit participant recall time and make it easier for participants to remember events and circumstances. Measures like expectations, intentions and attitudes may be prone to recall error as a result of “post hoc rationalization” [49, 12].

2.1.2 Disadvantages of longitudinal surveys

The first and most destructive disadvantage of longitudinal surveys, also known as their "Achilles heel", is sample attrition/panel attrition. Attrition is a major impediment towards attaining usable data [71]. Attrition is the continuous loss of participants during the course of a longitudinal survey due to non-response. It is however, important to note that in most cases a longitudinal survey will continue to attempt to contact the respondent in subsequent iterations, in most cases with success [23].

The second disadvantage of longitudinal surveys involves how they often succumb to panel conditioning. The way in which a first-time respondent and a repeat respondent fill out a survey are typically different. This is as a result of the conditioning that happens when a participant is exposed to previous conditions. This can come in two forms: the participant may have a change in behaviour, or they may have a change in the way they respond [35].

Cost is also a major disadvantage of longitudinal surveys. The initial and ongoing costs of longitudinal surveys are higher than that of other survey types. The work and planning required before the start of a longitudinal survey is greater than that required by a cross-sectional survey. A n-wave survey will have a greater cost than a once-off survey with the same sample size [14].

2.2 Survey design considerations

In order to design a survey that takes our project needs, such as monitoring levels of attrition, into account, this research needed to consider a variety of factors. Each is defined and justified in the sections that follow.

2.2.1 Sample population

Sample population is a necessary consideration in this particular study as different populations will give different results. Past research has shown that monetary compensation is less likely to have an effect on participation in affluent sample groups than in marginalised groups [32]. A sample population can either be static or cross-sectional depending on the goal of the survey. A static population means that the population size is fixed. Attrition is allowed but no new participants are

allowed to join the participant pool. Cross sectional populations, on the other hand, allow both attrition and the recruitment of new participants at each survey iteration [68].

A static population makes comparisons from wave to wave relatively easily, however, if attrition levels are high, data could become unusable. On the other hand, cross-sectional populations have the advantage of being able to keep a healthy sample size but as a result, this population could also introduce bias and makes analysis difficult.

2.2.2 Sample size

The aim of sampling approaches is to draw a representative sample from the population. This allows for the results of the study to be generalised. Generalisations is important as it allows for other researchers to take findings and apply them in multiple scenarios. In the case of our project, researchers will be able to identify the incentive that best suits their project needs based on our findings. There are two common ways to sample, randomly and statistically [52].

Random sampling of the population is ideally defined, and each member of that population has an equal chance of being selected for the study [52]. It is widely understood that random sampling is rarely used in field surveys as it is hard to achieve compared to statistical sampling [48]. Statistical sampling, on the other hand, involves the researcher using the population size, a confidence level, and a confidence interval, to calculate the desired sample size [8]. For this project statistical sampling was conducted as the goal was to identify the attrition rates of the four sample groups.

2.2.3 Geographical clustering

Clustering refers to researchers choosing participants in one localized area. Clustering is typically based on the number of survey waves, timing of waves, resources available, and richness of data required [68]. For example, if a longitudinal survey is being conducted to make decisions at a national level then geographical clustering would not be ideal as conclusions cannot be drawn from a survey run in one geographical location. For this project we chose not to use geographical clustering as we required our four sample groups to be in geographically separate locations in order to avoid sample contamination.

2.2.4 Mode of delivery

Surveys can be conducted over the web, over-the-phone, sent in the mail or face-to-face [26]. Web-based surveys are self-entry surveys that collect data over the Internet. Web-based surveys have the advantage of wide reach, low cost and the utilization of graphical user interfaces. However, web-based surveys also introduce coverage error, which in turn leads to a difficulty in sampling as well as non-response bias [3].

Mail surveys can either be sent by post or electronic mail. Surveys by post have often been criticized for their extremely high non-response bias which leads to unusable data [6]. While the response rates in e-mail surveys are higher than post, there has been a steady decrease of response rates since the inception of email in the late 1980s [65].

Over-the-phone or telephone surveys typically involve a verbal exchange or touch tone interaction. Over-the-phone surveys have provided researchers access to a large percentage of the world's population, however barriers such as voice mail, answering machines and caller ID technologies have made it difficult to prevent soft attrition in longitudinal surveys [24]. Telephone surveys are often considered to be an expensive alternative to web surveys primarily because they provide similar amounts of convenience [31].

Face-to-face surveys or in-person surveys are considered to be the costliest mode of delivery. This is true in most cases but can also be dependent on design considerations such as geographical clustering. In recent times telephone surveys have replaced in person surveys as the quality of data collected is comparable [4].

Modern researchers advocate for the use of mixed-mode data collection as it allows for the benefits of more than one survey collection method [26]. This study, however, chose to use the over the phone method as it allowed us to reach all our participants in a cost-effective manner at a time of their choosing.

2.2.5 Survey duration

Data that is collected over a long period of time is likely to be richer and more valuable, however, the longer the survey the costlier it is [49]. Survey duration can also be the difference between contacting your participants or not, as factors such as migration come into play the longer a longitudinal survey lasts [50]. In the case of our study, it is very important to have short intervals between surveys as seasonal work forces participants to move from farm to farm in search of jobs.

2.2.6 Survey length

While several studies have shown that survey length does not influence response [4, 25], other studies show that survey length has influenced response rate [65, 56]. Dillman et al. say that the length of the survey may not have an effect on isolation but, coupled with questionnaire design and question difficulty, could have an effect on response rates [25]. Muller et al. suggests leaving out "nice-to-know" questions from a survey in order to lessen the length and in turn reduce participant burden [56]. Singer describes telephone surveys that last less than an hour as low burden and describes survey length as a design factor that potentially has more of an effect on attrition than incentives [66]. We therefore attempted our best to keep surveys under an hour, by leaving out "nice to knows" in order to ensure that the burden to participants was as low as possible.

2.2.7 Survey compensation

Compensation for participation in under-served Africa is currently a hot topic as researchers have been describing reimbursement as "benefits", which have created tension between households and researchers [55]. A survey run in Zimbabwe revealed that participants of medical research were not receiving any compensation for their participation in research projects run in their community which caused controversy among the country's research fraternity. The survey revealed that it was an expectation of participants to be compensated for participation [53]. This survey showed why it is important for our research to explore compensation and incentives.

2.3 Attrition in longitudinal surveys

In this section attrition is examined in detail and under what circumstances attrition is highest, with a focus on both developed and developing regions. Attrition is formally defined as the action of dropping out of a study, refraining from picking up calls or verbally confirming the wish to withdraw from further participation at some point [2]. The rates of attrition in longitudinal surveys tend to be high due to the high participant burden and the possibility of participants losing interest over time [75].

According to the results from the British Household Panel Survey, the three main causes of attrition can be summarized as the failure to: locate participants, contact participants, and gain co-operation [32]. Other reasons such as participant fatigue and inappropriate length of survey are also known to lead to attrition [35].

Attrition in the developed and developing world tends to be different in general. In the developed world, refusal to participate is the main reason for attrition while in the developing world, constant migration leads to attrition [71]. It is however very important to note that there is a difference between attrition and non-response in the case of developing countries.

The difference between attrition and non-response lies in the participant's involvement. If a project participant verbally states that they no longer wish to take part in a project, then this is classified as attrition [36], which is true for most participants in the developed world. In the case that participants are difficult to track or do not respond, then this is classified as non-response [36] which is true in most cases in the developing world.

Combating non-response can be difficult with some studies going as far as asking participants for secondary contacts in the case that the primary participant migrates or is not reachable [71]. Couper and Ofstedal also suggest that having an alternative contact person is an effective way of battling non-response [19]. In this study, non-response is avoided by asking participants for the best time for us to call. This was seen to be effective due to the short space of time between survey iterations.

Harte et al. conducted a survey to rank the reasons for attrition. They found that personal motivation was the biggest issue in longitudinal studies [41]. Personal motivation of participants is very important in longitudinal surveys as displayed in Gouveia et al., where they narrate the adoption and engagement of an activity tracker called Habito. They found that the personal motivation of participants was low and therefore levels of attrition were high [38]. Personal motivation can be increased by the provision of direct feedback that is of interest to them as seen in many projects [21, 46, 35, 15, 21, 32]. In this study, we provide personalised direct feedback as a means of increasing motivation.

2.4 Factors affecting attrition

In order for our study to effectively measure attrition between the four sample groups it is necessary to look at the factors affecting attrition. We identify these factors to help minimise attrition across all sample groups. There are four major factors affecting attrition, namely, the ability to locate sample members, the ability to contact sample members, obtaining cooperation, and respondent characteristics [32]. Magruder and Nattrass agree with the above factors affecting attrition. They conducted panel surveys in Cape Town and found that general factors of attrition exist for shack-dwellers and small households [51].

Locating the sample member in most longitudinal surveys is the very first hurdle in attaining a respondent's participation. Surveys with a longer length of time between waves usually suffer from respondent migration [74]. Where movement has taken place, the tracking of movers is largely down to the tracking mechanisms implemented prior to the start of the survey [75].

Contacting the sample member is unlike trying to locate sample members as it is influenced by accessibility of dwellings in the case of face-to-face surveys and the use of an answering machine in the case of telephonic surveys. There are a number of survey designs that influence the successful contacting of participants and these include: number of calls attempted; timing of the call; number of attempted calls; number of survey waves and length of the study [75]. Nicoletti and Buck also established that interviewer continuity both at the contact and cooperation stages have an effect on survey response [57].

Obtaining cooperation is dependent on incentives in most cases [49]. Longitudinal studies have been known to use incentives in the form of money or gifts in order to reduce attrition. The effects of incentives have been known to vary depending on application and administration [46]. The interviewer effect can also be integral in obtaining cooperation. The interviewer has a vital role to play in gaining the cooperation of participants [49].

Mode of delivery of surveys is also a factor affecting attrition. Singer et al. conducted a review to assess whether the mode of delivery, incentive and provision before or after had an effect on participation. They found that there was a significant difference between mail surveys, telephone and face-to-face survey, due to high burden in mail surveys. They also found that there was no relationship between burden and incentive [66].

Respondent characteristics are also an important factor. Literature tell us that women are more likely to respond to survey calls, the main reason being that women are at home more [57]. It is also widely reported that response rates are lower among the young population, but result are relatively unclear for the older population. Married people are more likely to agree to respond to surveys than single people. Expectedly, employed people are more likely to drop out of a study than unemployed people. The effect of race and ethnicity on response rate is unclear [75].

Subject matter is not a major factor, according to Fumagalli, however Dillman believes it is an important factor. Dillman suggests that asking difficult questions may be the difference between a response and no response [25], however, other research suggests it is the way that you ask questions that has an effect on response [29].

2.5 Strategies for minimizing attrition

Minimising attrition is a priority for all researchers conducting longitudinal surveys. Incentives are often used as a means of increasing participant motivation for continual participation in a study [46]. For this study minimising attrition meant (1) the strategy that works best could be measured (2) the sample size stayed large enough for statistical analysis, (3) results could potentially be generalised. This section identifies strategies that are currently used to minimise attrition and how they are used.

2.5.1 Monetary incentives

In general, researchers provide an incentive for survey participation. Incentives typically consist of monetary compensation or gifts. Past research has shown that participants prefer to be compensated for their time [76].

Grady discusses four compensation models used to establish suitable compensation for project participants. The first is the market model, which serves as an incentive and the amount paid out is driven by supply and demand of participants or “market rates". The second is the wage-payment model, which serves as compensation for a service rendered. The reimbursement model simply pays for participant expenses. Grady also speaks of the appreciation model, which serves as a reward for work done and is usually given at the end of the project [39].

Laurie and Lynn suggest the use of incentives such as gifts and money, not only as a tool to reduce the level of attrition but also as a tool that boosts the confidence of the interviewer, as they have something to give the respondents. Having a confident interviewer is positive and can lead to increased productivity [82]. Laurie and Lynn also found that the form of incentive and the way in which they are delivered both have measurable effects on attrition. They claim that money has a greater effect than gifts on attrition. In addition, due to the increased participant burden seen in longitudinal surveys, money is more likely to be used. Laurie and Lynn also suggest, monetary incentives given before the survey have the greatest effect as they act on trust and the concept of social reciprocity [46].

While monetary compensation has shown to be effective in some cases, in other cases it has been identified as problematic. This was the case in an m-health project conducted in Malawi. The project’s data collection was intended to be done by local community health workers, however, their superiors had taken over that responsibility so as to gain the monetary compensation for themselves [63].

There is other research that suggests alternative “hands-on” methods to reducing attrition due to the negatives that result from monetary compensation [19]. This is mainly due to the negative impact that monetary compensation can have on respondents as demonstrated by Gerken. Gerken’s study on longitudinal research found that there is a danger in using monetary incentives as they

create bias [35]. The other drawback of monetary incentives is the expectation participants build. Once a participant has been monetarily incentivized, they are unlikely to participate in a future survey for anything less [73, 46].

2.5.2 Alternative methods for reducing attrition

A known strategy used to minimise attrition is participant tracking. For telephone interviews in particular, past research suggests the following to minimise attrition: trying all available numbers for each participant, checking current listings for changes in codes (area/city/regional), checking with directories or operators for any phone number changes, calling a designated secondary contact and sending out a letter/email to respondents requesting new phone number [19].

Another strategy involves observing the culture of your sample group. Some social science scholars argue about the effects of "doing well vs doing good". This is down to how a person feels about their social image. Ariely states that in some cultures, people are motivated to do good when they are not offered a monetary incentive due to the image they gain socially and therefore the offering of monetary compensation lessens motivation [5]. It is also known that in some cultures praise and increased reputation can also work toward lowering the rate of attrition [81].

Zimmerman et al. found that entertainment and/or learnings of interest can be used as a method to reduce attrition [34]. In our project we used this notion as we provided some sample groups with feedback that we felt would be of interest to the communities with the assistance of enumerators that were part of the community. Research like Zimmerman et al's compliments this project because it uses feedback as an incentive for continued survey participation [21, 46, 35, 15, 21, 32]. Furthermore, we found that in most clinical trials, feedback is used as a mechanism to reduce the rate of attrition with great effect [20]. From a business and economic stand point, performance feedback has long been used to motivate employees [45].

2.6 Feedback as an incentive

Our project defines feedback as informative information delivered to survey participants by the researcher. Feedback in this context has three key features, (1) it is based on the answers given by each individual participant, (2) it must be short and simple enough for the participant to understand,

(3) it should be educational, informative and of interest. This section discusses how other projects have viewed and used feedback.

Feedback is used to increase awareness and motivation of participants [46, 15]. Some researchers also use feedback as a means of increasing the strength of the bond between researcher and participant as a result of the added communication [35]. The method of delivery when giving feedback is very important and there are two factors that should be considered when choosing the technology: cost of implementation and impact of the feedback [37].

Feedback is known to be effective when it is given frequently and therefore, delivery methods have to allow for frequent updates [37]. This also means that survey iterations may need to be shorter for feedback to have an effect on participant motivation.

Most of the literature on providing direct feedback to participants as an incentive is in clinical trials, although there is some early human computer interaction work that mentions the concept [35]. Card et al. for example, conducted a study in which four devices were evaluated to establish how rapidly they could be used to select text on a CRT display. They found that feedback played an important role in participant motivation as their experiments had a large burden on participants [15]. Similarly, a survey by Cox et al. discovered most research participants that had not received feedback were “overwhelmingly” in favour of receiving feedback, but researchers did not often provide it. This was mainly due to the perceived effort involved with providing direct feedback to participants [20].

2.6.1 Effects of feedback in the field

In developed regions, there have been a few studies done in which participants receive some form of feedback via text and email. This choice might be because SMS and email are cheaper to the calling alternative. Gleerup et al. conducted a study that assessed the effects of feedback on electricity consumption. They found that feedback helped in motivating households to participate [37].

Fumagalli et al., on the other hand, experimented with providing participants with feedback inside a brochure. This was conducted in an attempt to get participants interested in the project and ultimately reduce the rate of attrition. They found that the brochures encouraged people that previously would not have participated to answer the telephonic surveys [32].

Pedersen et al. worked on a project that delivered feedback to adolescents on their fruit and vegetable intake. The project found that the use of SMS for feedback and interactions with the teenagers affected the level of engagement from the teenagers and they noticed more active involvement as a direct result of the feedback [59].

A very similar study, specifically aimed at tackling child obesity, used feedback via SMS to change behaviour in overweight children. This study was conducted by Fassnacht et al [28]. The study showed that feedback, specifically via SMS, resulted in the motivation of participants to change their eating and exercise patterns regardless of the fact that no further incentives were provided. Their system delivered tailored feedback based on participants responses. The study concluded that the response rate when using SMS was much higher than that of previously used paper-based surveys, which was attributed to any of the following reasons: the children's adherence to technologically enhanced systems, tailored feedback and the reduced need for self-monitoring [28].

2.7 Survey technology

In order to conduct surveys, there are some technological considerations that had to be assessed. The study surveys involved people that were to be contacted at different times of the day and lived in separate geographical locations. This section describes which survey technologies were available and how they would be used to serve the project.

Collection of Data is a very important consideration when choosing survey technologies. Data collection today is conducted using web-based or phone-based platforms. Web-based surveys are usually cheap to conduct, are easy to use for both researcher and participant, and have been found to lead to more honesty on the part of participants. Phone-based surveys, on the other hand, are fast, easily accessible, and often do not require Internet access [1].

Despite the mentioned advantages, web-based surveys suffer from their need for an Internet connection which, in under-served regions, renders them inaccessible [78]. Mobile surveys, like web-based surveys, also have disadvantages such as their predisposition for android OS and the fact that mobile devices have finite amounts of disk space [1]. This project used phone-based surveys as they do not require an internet connection and reduce the burden on participants.

2.7.1 Web-based technology

Web-based survey technologies have become very similar in recent times and therefore, if one were chosen for this project, it more than likely would have been one that had everything needed for the study. To illustrate their similarity, five popular platforms were chosen, namely, SurveyGizmo, Form site, Polldaddy, Question pro and Survey Monkey. All the aforementioned platforms offer free survey creation service. All of them can be used commercially. All allow you to add design to your survey and allow for multi-lingual support. All but one have social media integration. All have in-build data analysis tools. And all of them offer data sharing [1].

Web-based survey technologies allow for participants to take surveys by themselves in their own time. Whilst this may have been valuable in the battle against attrition, in the case of our project, our target demographics were unlikely to have computers with Internet connections and therefore web-based surveys were not viable.

2.7.2 Phone-based technology

Phone-based technologies are very different to web-based as they are made for a different context. Typically, phone-based surveys are designed to cater for data collection by enumerators whereas web-based surveys follow a self-service model. Furthermore, phone-based technologies usually allow for offline data collection, which in turn allows for data collection in remote areas. The following are four popular mobile survey tools: (1) Open-Data-Kit (ODK) is an open source data collection tool used in mobile data collection. ODK uses survey forms on android devices to enable data collection [42]. (2) OpenXdata is also an open source platform that allows for design of surveys, management and data collection on low cost devices [80]. (3) Magpi, formally known as Episurveyor, is a web-based survey platform for mobile phones. Magpi offers offline support via

the use of SMS surveys [1]. (4) FrontlineSMS is an open source software used to distribute and collect information via text messages (SMS) [9].

For our project, phone-based data collection was used as it allowed for offline data collection, which was used in many instances as explained in great detail in the dissertation. In particular, we chose to use Open-Data-Kit due to our prior knowledge of the technology and the readily available android devices that we could use for project purposes.

2.8 Case studies

In the following section, four case studies that did similar work to our project are detailed. The goals and takeaways of the studies are analysed, so as to compare their findings to ours at the end of our study.

2.8.1 Cutting the costs of attrition: Results from the Indonesia Family Life Survey

Thomas et al. accessed the success of a longitudinal survey that was conducted over a duration of fourteen years. The survey collected information on health, ran an average of 39000 surveys over four iterations, and gave participants US$3 as a “token of appreciation”. In the methodology of the project they mention that they planned to minimize attrition and non-response by “careful attention to planning and training, persistence and commitment to the study, assuring the co-operation of all respondents and judicious allocation of resources”. This plan in its entirety is one that was kept in mind at all stages of our project as the four aforementioned considerations were the reason for the low attrition rates in their project. They had relatively low attrition in comparison to similar projects and were therefore relatively successful as shown in table 1 [71]. The table shows the project’s attrition levels over the 14 years which were all below 2.5%. This leads us to conclude that Thomas et al’s plan to minimise attrition was successful.

Table 1: Attrition and Non-response in an Indonesian 14-year longitudinal survey

	1993	1997	2000	2007
Attrition	2.5%	1.1%	0.8%	1.1%
Non-response	0%	8.6%	8.5%	13.4%

Thomas et al. did, however, find that locating participants was significantly more difficult to do with time. They also deduced that the US$3 was an effective incentive. These findings go against the social science theory that claims that money can be a driver of attrition [5]. Thomas et al. also states that these numbers were achieved by implementing strict follow-up rules that included participants identifying individuals that would know of their whereabouts if they were to move from their current locations, which is recommended by Lynn [49]. Participants were essentially paid to take a health check-up, which was another contributing factor of the low rates of attrition [71].

2.8.2 Challenges and opportunities of mobile phone-based data collection: Evidence from South Sudan

On the other end of the attrition scale, due to its high levels of attrition when compared to the case study in section 2.8.1, is a project that was conducted by Demombynes et al. The project collected mobile data in South Sudan. All the participants in this project where provided with incentives in the form of a mobile phone and mobile credit. Participants were split into four groups, one received US$2.17 and a either a Nokia (Group N2) or Safaricom (Group S2) mobile device. The others received US$4.35 and a either a Nokia (Group N4) or Safaricom (Group S4) mobile device. The project also offered an additional incentive in the form of a monthly draw with a grand price of US$100, granted the participant completed their survey. Survey calls were delivered from a call centre and surveys lasted 15-20 minutes. Data collection was conducted monthly for four months.

Attrition was relatively high during the course of this project with only 31% of participants completing all four surveys and 17% failing to complete a single survey [22]. The project saw contradicting results as Group N4 had higher attrition than Group N2, whereas Group S2 had higher attrition than group S4. These results showed how the provision of more money may not

necessarily equate to less attrition. The results also indicated that monetary incentives are not always an effective means of reducing attrition. This in turn, shows the importance of research like ours. Is there an incentive that could potentially be more effective at reducing attrition that monetary incentives?

2.8.3 mClerk: enabling mobile crowdsourcing in developing regions

Monetary compensation is not a "silver bullet" for the problem of attrition [35, 63]. This was evident in a longitudinal crowd-sourcing project called mClerk. MClerk was established to assess the potential for crowd-sources transliterations of words from Kannada to English. The project was run over a five-week period and was left to grow organically from ten initial participants all the way to 239 participants.

Participants were compensated monetarily with the value of compensation changing once during the project. The project began with a participant compensation of *INR*0.5 per word correctly transliterated. The project then slashed the compensation to *INR*0.2 which caused an increase in attrition. This compensation method proved fatal as the project saw attrition levels of up to 53% between the two iterations. This particular incident showed the need for a sustainable incentive like feedback.

Interesting to note was that participants that joined the project after the compensation had been slashed to *INR*0.2 did not drop out until the end of the project [40]. This may have been due to the fact that they had not experienced the higher compensation. The experiences seen in this project show that monetary compensation can be a motivating factor in participation, however once taken away, they have negative effects on attrition rates. This is also an example of the negative effects of monetary compensation as they can build up an expectation within participants [73, 46].

2.8.4 Estimating the effect of incentives on mail survey response rates

A meta-analysis study that involved finding the effects of monetary and non-monetary incentives analysed four scenarios that consisted of monetary and non-monetary incentives mailed with the survey (MI and NI respectively) and monetary and non-monetary incentives given on the return of the survey (MR and NR respectively).

The study found that MI was significantly different to all other incentive types. This showed that participants were most motivated to participate when they received a monetary incentive with the survey. The study also found that there was a significant difference between NI and NR. Participants preferred to receive a non-monetary incentive before participation. There was however, no significance between NI and MR. The results showed the strength of providing participants with their incentive before the survey as opposed to after the survey [18]. Our study, however, has chosen to give the incentive after the survey as feedback cannot be constructed before the survey has been taken.

3 Study design

In order to answer our research questions, we designed a study to assess the effect of feedback on attrition in comparison to monetary incentives. It was also pertinent to look at the costs of providing feedback as literature describes that high costs usually deter researchers from the provision of feedback [20].

This study was designed to explore and answer the following questions:

1. How do different incentives affect attrition?
2. How do different incentives relate to a participant's habits during a longitudinal survey?
3. How do different incentives affect the cost of a project?

Therefore, a one factor study was designed between subjects, with four levels. Our factor was attrition and the four levels consisted of four treatments. The treatments consisted of four groups of participants located in separate locations. Each treatment was almost equal in size and received a different incentive compared to other groups.

The design of this study allowed a comparison among the provision of mobile credit (the airtime group), feedback delivered by the enumerators (the voice group), feedback delivered by an automated feedback delivery system (the automated group) and a control that did not get an incentive (the control group).

It is reported that researchers do not often provide feedback to participants due to the perceived effort associated with doing so [20]. The study, therefore, altered the feedback delivery mechanism so as to assess the effectiveness, effort, and cost of recorded feedback compared to the interaction between an enumerator and a respondent.

Thus, this experiment required a number of design considerations. Section 2.2 identified seven major issues that needed to be considered when running a longitudinal study in under-served communities, which are each described in the sections that follow.

3.1 Design Considerations

Longitudinal surveys face a few sample design issues that are highly dependent on the study design. Section 2.2 discussed how decisions such as dealing with sample changes, survey length, number of iterations and geographical clustering, are important for the success of a longitudinal study.

3.1.1 Study sample size

Sample size was mainly determined by the resources available and an acceptable confidence interval. It was found that with roughly 11,000 households in the four locations, if 200 households (50 households per location) were recruited, using the worst-case percentage of 50% and a confidence level of 95%, there would be a confidence interval (margin of error) of 6.87%. The project budget also helped determine the number of participants that could be recruited as there was a finite amount of resources.

3.1.2 Geographical clustering

The study required four locations, which translated into four groups, each of which received a different incentive for participation. These locations were demographically similar in order to allow for comparison but had to be geographically separate in order to reduce the risk of contamination. In this study, contamination bias was identified as a potential for any of our group members meeting. Therefore, it was rationalized that no geographical clustering was needed for this study.

As discussed in section 2.2.3, geographical clustering involves choosing participants in one localised location. The need for geographically distant locations meant that the surveys could not be conducted in-person due to the costs and time that are associated with in-person surveys.

Contamination bias was also an attribute that was taken into consideration when determining whether geographical clustering was ideal [62].

3.1.3 Study duration and intervals

Due to the high recall data that was to be collected, the project was run for a twelve-week period with very short intervals. Surveys and feedback were run bi-weekly for the twelve weeks. Therefore, there were six surveys and six feedback rounds conducted. The short nature of the study and frequency of the intervals increased the response burden of the respondents and therefore higher attrition rates were expected to be observed than seen in previous studies [75]. This would also allow the researchers to find if there was any significant difference among all incentives.

Survey duration is particularly important because of the quality of data and the cost of running the survey are proportional. Survey duration is also directly linked to the length of intervals between waves. Intervals are dependent on study design and the type of data being collected. High recall data will require high frequency for high quality data. As a rule of thumb, the frequency needs to be low enough to capture change but not too long as events may be missed [35].

3.1.4 Mode of delivery

For this study over-the-phone surveys were preferred because of reduced cost and convenience mainly due to the geographical distance between our locations. By extension, over-the-phone surveys also allowed for the channelling of resources towards incentives due to their low cost nature. By design, over-the-phone surveys enabled us to ask participants what time they preferred to receive calls for surveys. The other ways to conduct a survey are: face-to-face and self-completion [49] as described in section 2.2.4.

3.1.5 Survey length

The project aimed to make sure that all surveys were at most 30 questions and/or 20 minutes long. Singer describes telephone surveys that last less than an hour as low burden [66]. In addition, previous studies have shown a correlation between survey length and non-response [60]. As discussed in section 2.2.5, there is evidence that survey length is related to response rate and by extension attrition. Therefore, to keep the survey length factors effect on attrition minimal a

consistent survey length was used. While over-the-phone surveys have the advantages of convenient, high-quality data with low error rates, they increase the difficulty of contacting participants [68].

3.1.6 Survey compensation

A hybrid model was used to compensate participants. The market and appreciation model were used to compensate participants as the aim was to reward participants for their participation using the current remuneration rates in the country. This also meant that participants were provided with the incentive after their participation.

It was estimated that a survey would take anywhere between 15 and 20 minutes to complete and based on that knowledge, the country's working wage was used to come to a rate of *R*10 for each completed survey. As this research did not involve participants' personal funds, no reimbursement was necessary.

3.2 Survey development

The development of surveys was done bi-weekly using an iterative process. ACDI played the role of the subject matter experts and therefore were responsible for the development of surveys. The six surveys asked questions on: general well-being, water use and water availability, food security, energy security, well-being in the context of water, energy consumption, the natural environment, and extreme events in the area/town. These topics were very relevant as the areas in which the surveys were being conducted were currently going through a water crisis due to drought.

3.2.1 Survey development process

The survey deployment process aimed to refine the survey questions in order to eliminate offense or intrusiveness. We established an 8-step process in the development of surveys:

Step 1. First draft of survey established by ACDI
Step 2. Survey reviewed by researcher and sent to ACDI with comments and suggestions

Step 3. Survey updated by ACDI and sent to researcher.
Step 4. Researcher forwards survey to FLOW, survey is reviewed and returned.
Step 5. Researcher forwards survey to ACDI for final review.
Step 6. Final draft is established by ACDI and sent back to FLOW.
Step 7. FLOW translates survey to Afrikaans.
Step 8. Researcher codes survey into Open Data Kit.

3.2.2 Survey deployment process

The main goal of the deployment process was to establish an actual survey length in minutes and appropriateness for the survey. This process was conducted for every survey cycle. The 4-step process involved:

Step 1. Face-to-face timed English interviews with dummy participants
Step 2. Face-to-face timed English test interviews with enumerators
Step 3. Face-to-face timed Afrikaans test interviews with the use of ODK
Step 4. Over-the-phone demos. Enumerators call FLOW coordinators

3.3 Feedback development

In this study, feedback was given to participants with the goal of encouraging participants to continue their involvement in the project. The development and delivery of feedback was therefore done bi-weekly, the week after a survey was conducted. An agile iterative process was utilized in the development of the feedback. Feedback was made up of five parts: introduction and thanks; statistical overview of the community's current state; constructive praise/encouragement based on the answers they provided in the surveys; some general advice on improvement of lifestyle for the community; conclusion and thanks.

3.3.1 Feedback development process

In order to keep the feedback consistent a process that would be followed was established for each feedback iteration. The six-step process consisted of:

Step 1. Analysis of survey data and generation of descriptive and summary statistics
Step 2. Develop feedback draft and forward to ACDI
Step 3. ACDI review and send back comments of the feedback
Step 4. Feedback updated and then sent to FLOW for further comments
Step 5. FLOW reviews feedback and returns comments
Step 6. Final feedback draft is established and sent to FLOW for translation

After translation, feedback would be revised with the enumerators in English and Afrikaans in order to get their comments and for the researcher to provide explanations.

3.3.2 Feedback recording (for automated group)

The recordings were conducted by the same enumerator at the end of the feedback development process. The feedback was recorded on a laptop using a set of standard earphones with a mouth piece, after-which Audacity [70] was used to clean and amplify the audio files.

3.4 Selection of towns

Together with ACDI, it was established that the study would target towns in the Bergrivier Municipality due to FLOW being located there. The choice of town in this particular location was based on statistics recovered from Statistics South Africa. They showed great similarities in the demographics of four towns in particular. The four chosen towns were Piketberg (Airtime group), Veldriff (Voice group), Porterville (Automated group), Moorreesburg (Control group). All towns but Moorreesburg are in the Bergrivier Municipality. Moorreesburg is located 20 kilometers outside the Bergrivier Municipality boundary lines and contains similar demographics as the other towns.

3.4.1 Demographics of towns

Due to the project design, the demographics of the towns had to be similar so as to allow for comparison. As shown in table 2, the formal dwellings in all four towns are between 89% and 93%. Average household size is also very similar with all the towns, between 2.9 and 3.8 people per household.

According to statistics South Africa [83] the demographics of all four locations are as follows:

Table 2: Town demographics of Piketberg, Porterville, Moorreesburg and Velddrif

	Occupants	Households	Formal dwellings	Refuse removal	Avg household size
Piketberg	12075	2920	90.6%	96.9%	3.8
Porterville	7057	1949	93%	99.4%	3.5
Moorreesburg	7760	2578	93.6%	97.9%	2.9
Velddrif	11017	3622	89.5%	98.5%	2.9

3.5 Recruitment of enumerators

Using FLOW's knowledge from their work collecting data, it was established that the study required four enumerators to complete the required tasks in the allocated time. The selected enumerators would be required to have the ability to speak and write well in Afrikaans and English as well as be computer literate. They also needed to be available a few days a week. They needed to be fluent in both Afrikaans and English in order to clearly communicate with participants and researchers respectively.

The FLOW co-coordinator made the final decision on which enumerators would be suitable as he had in-depth knowledge of the enumerator's capabilities. He chose three enumerators out of the 2017 stream of FLOW ambassadors and one from the 2016 stream. The three enumerators from the 2017 stream had never conducted such work and therefore the forth enumerator was brought in as a quality control measure as she had done similar work for FLOW and the Bergrivier Municipality. Three of the four selected enumerators were from Piketberg (Airtime group), including the enumerator from the 2016 stream and the last was from Porterville (Automated group).

3.5.1 Recruitment training

Recruitment training methods detailed by Burke et al. were used for worker health and safety training. Training in health and safety is crucial as it is the first defensive in avoiding accidents. They recommended using practice and dialogue for effective training [13].

All four enumerators were present for training, as well as one FLOW ambassador supervisor and four interns from ACDI. The training started with a short icebreaker. Icebreakers have been shown to have positive effects in spaces where people aren't familiar [17]. To further keep the enumerators relaxed, the training was designed as informally as possible, which would in-turn allow the enumerators to participate in the training session [16]. After the icebreaker, the project was introduced and talk about the goals of the project, both from the computer science and the ACDI perspective, and ran through some recruitment and survey guidelines. The guidelines included: courtesy, introductions, professionalism, consistency, speech clarity and politeness.

We then reviewed a sample recruitment survey and a demo of the recruitment process was conducted. After the demo, there was a discussion with the enumerators on their thoughts regarding the recruitment process. This marked the end of the morning session and during a refreshment break there were informal discussions.

After the break, the enumerators went through the recruitment process, which was made up of: reading the project summary to participants, asking the participants to sign consent sheet, conducting the recruitment survey using ODK on the devices they were provided, getting an electronic signature on ODK, and finally leaving the project participants with copy of the project summary.

Upon finishing, the enumerators underwent demo recruitments. They recruited members of the ACDI team. The normal recruitment environment was simulated by having them stand on one side of an office door while the "participant" stood on the other side of the door. The enumerator knocked on the door and thus began the recruitment process. They had to recruit in both English and Afrikaans as it was known that some of the towns might have English speaking participants. This training method kept the participants attentive and engaged. As a result, great knowledge acquisition was observed similar to that seen in Burke et al.'s project due to the interactive training methods [13].

3.5.2 Outcomes of recruitment training sessions

During the training sessions three details were discovered that were initially overlooked. Firstly, the experienced enumerator suggested that the researcher get badges for the enumerators as it

would help with their legitimacy in the eyes of the communities. It was also found that even though the enumerators understood English, their participation seemed to be limited as they were not completely confident speaking English. Furthermore, it was found that the enumerators should ask whether a participant had a mobile phone at the beginning of the recruitment survey as some potential participants would be ineligible due to lack of a mobile phone.

3.6 Assigning incentives

In order to establish incentives per location, the researchers sat with the partners from FLOW to discuss the project goals and what was required. In choosing treatments, Piketberg was chosen to receive airtime under the assumption that if Piketberg was the location to receive airtime, people were likely to readily participate in the study, which in turn would boost the enumerators' confidence. This assumption was also backed by Laurie and Lynn who say that the provision of an incentive is often accompanied by the hope that it will raise interest in participation particularly by groups that would not have participated normally [46]. Furthermore, it was known that the local municipality had run a survey in the community and therefore felt that participants would need an incentive to make them receptive to recruitment.

Further incentives were assigned to the other towns. However, on second thought, we realised we were meant to randomly assign incentives. We subsequently ran a randomization script in R that would indicate how each location would be incentivized. We assigned each of the treatment groups to a number between 1 and 4 as follows: 1 - Piketberg; 2 - Porterville; 3 - Velddrif; 4 - Moorreesberg. We then used R to randomly generate a sample to randomize the groups. The randomization indicated that we were to provide the airtime incentive in the first location, make automated calls in the seconds recruitment location, have the third location as the control group and conduct live enumerator calls in the fourth location. The results from the randomisation matched the pre-determined order assigned to the incentives.

3.6.1 Study ethics

Due to the nature of this study, close attention had to be paid to the ethics. This study therefore applied for and was awarded ethical clearance by the University of Cape Town. The study required all four groups of participants to be in different towns and for the participants to never come in

contact. This meant that during the recruitment of the participants, some level of deception had to be employed. As a result, only part of the study could be communicated. To counter this ethical dilemma, we revealed to the groups that the study was being conducted in various locations in the workshops conducted at the end of the study.

As stated above, each group received different incentives for their participation in the study and therefore at the end of the study it was of the utmost importance that all groups were put on an equal standpoint. This, for example, meant giving the control group both the airtime and the feedback that the other groups had received.

3.7 Recruitment of groups

With the knowledge that FLOW had previously conducted data collection, we chose to follow a similar plan to the one which had been used when mapping local businesses in Piketberg. The enumerators performed the mapping in pairs for their own safety as the enumerators were walking around with tablet devices.

Due to our strict recruitment criteria, we could not systematically pick our door to door recruitment as encouraged by previous studies. Previous studies typically choose to recruit every nth house when doing door to door recruitment [7]. Our inclusion criteria included households that received social grants but also had some additional source of income such as seasonal work or a part-time job. We therefore chose to start at the beginning of a neighbourhood that contained our desired sample and conduct door to door recruitment until we achieved the number of participants required.

To prepare for the recruitment of the households, we held a training session for the enumerators over two days, on a Thursday and Friday with recruitment scheduled to begin on the following Monday. Recruitment of towns was done over three days; a day longer than planned. We recruited the Airtime group on the first day, the Automated group on second day and on the third day households in the Control group and Voice group were recruited. About 50 households in each location were recruited in total.

3.7.1 Recruitment survey

The recruitment survey was short and to the point so as to save time during recruitment. As mentioned in the technology section, Open Data Kit was used to collect the participants' recruitment data. The questions asked during recruitment were as follows:

1. What is your first name?
2. What is your surname?
3. What town do you live in?
4. What is your home language?
5. Do you have a personal mobile phone?
    a. No, I do not have access to one.
    b. No, but share with the household.
    c. Yes. I have my own mobile device.
6. What is your phone number?
7. What is the mobile operator?
8. Electronic Signature

4 Technology and system design

This study relied on technology to achieve all survey and feedback aspects. In order to answer the three research questions two systems were built:

1. An automated voice call system (AVCS): designed to call participants and deliver recorded feedback.
2. A call tracking system (CTS): designed to keep details of enumerator calls to participants.

This chapter will introduce and describe the systems that were built for the project and the methodologies used in their development. The technology used in calls and the sending of mass messages are also described in order to give context to the analysis of project cost at the end of the dissertation.

4.1 System development

To develop the AVCS and the CTS we utilized an agile methodology. We identified the agile methodology as ideal as we wanted the ability to continuously re-design and test if there were usability or requirement changes.

4.1.1 System stakeholders

Three system stakeholders were identified as the enumerators, the respondents and the researcher. The enumerators are stakeholders as the users of the system, the respondents as the people that will receive automated feedback based on their survey answers, and the researcher as the stakeholders that develops the system, analyse the data and develop the feedback.

4.1.2 Requirements gathering

Prototyping was used as the main requirements gathering tool. A four-step process was used, which involved identifying initial requirements, developing a prototype, reviewing and revising the

requirements [67]. The final requirements that were uncovered from prototyping are described below.

4.1.2.1 Automated voice call system

AVCS was required because the study needed a system that automatically delivered feedback recordings to participants that answered surveys. We, therefore, identified that the system needed to be able to (1) communicate with CTS to retrieve the list of participants that had completed surveys, (2) call multiple participants simultaneously, (3) communicate with Google's API in order to download completed surveys, (4) deliver a recording based on the survey answers of the participants.

4.1.2.2 Call tracking system

CTS was needed because it was necessary to track calls throughout the projects. It was identified that enumerators would need to (1) identify participants they needed to call, (2) identify what time participants would like to be called, (3) identify participants that have already been called, (4) identify the state of the calls that had been made.

This functionality was also needed so as to enable the researchers to keep track of the enumerators' work. As such, CTS also had the functionality to download Excel spreadsheets that contained all the call logs and call states.

4.2 System design

4.2.1 Automated voice call system

The goal of the system was to enable us to make automated calls to participants in the automated group. The system programmatically used Twilio's API to make the calls. Twilio is a cloud communications platform that allows for developers to add voice, video, and messaging to their applications [69].

As seen in FIGURE 1, the AVCS was designed to programmatically gain access to our Google Drive folder and pull survey results using Google Drive's API. After retrieval of survey results,

the data is dumped in the AVCS's database. This allowed the system to have information on all the participants that answered surveyed and were to receive feedback.

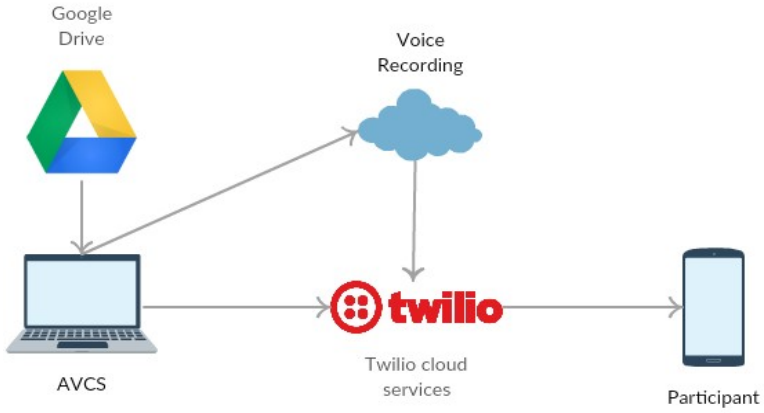

Figure 1: How the study's AVCS works

4.2.2 Call tracking system

CTS was designed to assist the enumerators to keep track of the state of calls, in other words, whether a call was completed, not completed, there was no answer, if the participant asked to be called later or if they asked to drop out of the study. The initial screen contained a list of all the participants to be called. The participant list was arranged in order of the time the participants indicated as an ideal call time. A screenshot of the of the CTS home screen is seen in FIGURE 2.

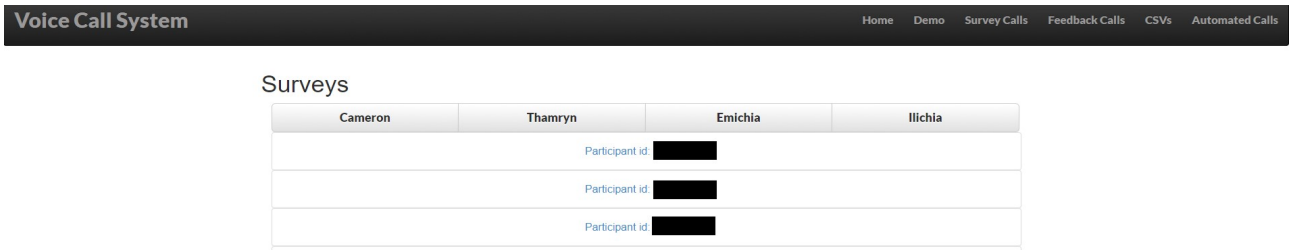

Figure 2: Participant list arranged by time

Once the enumerators selected one of the participants, the panel dropped down (shown in FIGURE 3) and showed some information on the participant. This gave the enumerators access to the participant's number, location and preferred time of call. Also available where the 6 call states. A

call could be answered and completed (ANC), answered and not completed (ANNC), not answered (NA), requested for another time (CAT), requested for the next round (CNR) or a participant could have asked to never be called again (NCA).

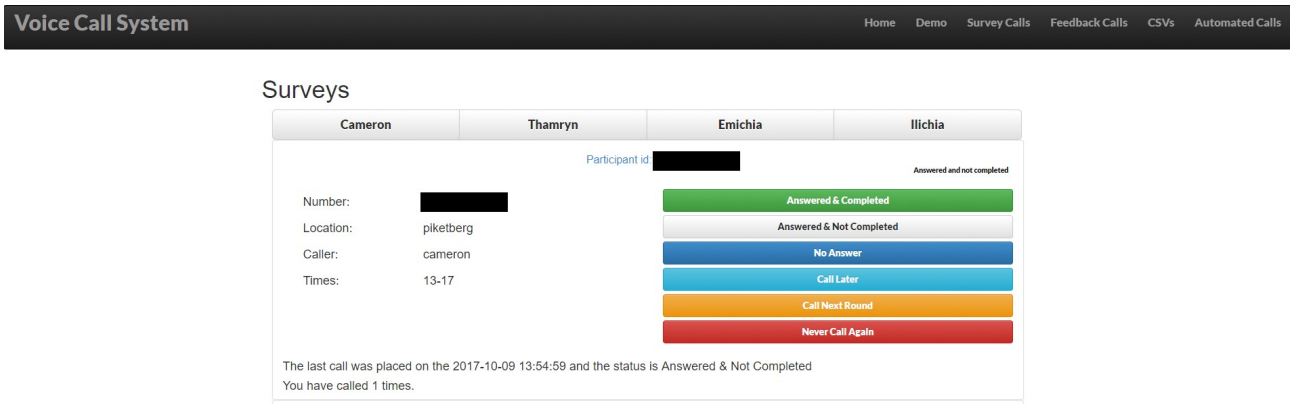

Figure 3: Open participant panel showing participant information and call status

Each of the buttons have different outcomes—Answered and completed, call next round, and never call again— These buttons all result in the contact disappearing from the call list from the current round of calls. The other three buttons, answered and not completed, no answer, and call later all result in a status notification on the participant panels, as seen in FIGURE 4.

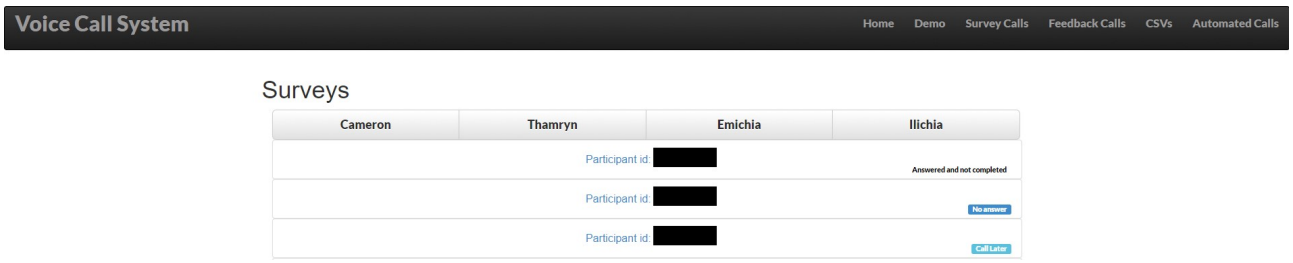

Figure 4: Participant list after attempted calls

The system was also designed to produce a downloadable Excel spreadsheet with all the outcomes from the enumerator calls. This allowed download of data for statistical purposes, more specifically, with the spreadsheet the following could be gathered: (1) the time enumerators made calls to participants, (2) how many times an enumerator called a specific participant, (3) what happened on each call. Table 3 is an example of what the downloaded spreadsheet looked like. It contained the participant ID, their phone number, the enumerator that called them, the time they

requested to be called, when the call occurred, the state of the call and how many times the enumerator attempted to call them before they answered.

Table 3: Sample system output

ID	Phone #	Group	Enumerator	Call Time	D/T	Status	Attempts
1	001	Airtime	E1	8-10	7-Sept 8:45	NA	1
2	002	Airtime	E2	8-10	7-Sept 8:46	ANC	2
3	003	Airtime	E1	8-10	7-Sept 8:47	CNR	1
4	004	Airtime	E4	8-10	7-Sept 8:48	ANNC	3

4.3 System user and integration testing

In testing the CTS, User Acceptance Tests [33] were predominantly used to determine whether the system was at a usability level that was acceptable to the enumerators. For the AVCS, we conducted integration tests [79] (due to the use of different APIs used) and usability tests.

4.3.1 AVCS testing

Usability and integration testing for the AVCS involved verifying that recordings were being delivered. To achieve this, the system was tested to see if it was making multiple calls simultaneously while delivering the correct feedback to the participants. To test that the system was behaving as described some phones numbers of volunteer postgraduate students were hard-coded into our system database, assigned one of four feedback recordings, and then the AVCS was run. The system operated as expected.

4.3.2 CTS testing

To test CTS, over-the-shoulder observations of the enumerators interacting with the system were conducted. Due to the system being extremely simple, it was seen that enumerators were able to log their call events with no problems. They were also able to accurately keep track of the state of the call. This was checked by downloading an Excel spreadsheet with all events on the CTS.

4.3.3 AVCS and CTS integration testing

In order to verify that the AVCS and the CTS were correctly integrated, end-to-end testing was conducted. In order to test this, dummy participants from the four treatment groups were gathered, completed surveys for each of them were created, and four different types of recorded feedback were made. Then the CTS was used to enter call states for all the participants after which the AVCS was run. The test revealed that only the participants that were in the automated group and had states of ANC were called.

The last phase of testing was conducted with the enumerators and was a live end-to-end test. A set of steps was drawn up that the enumerators had to follow when making survey calls and interacting with the system. CTS and ODK were both used on a tablet device and the calls were to be made on an android mobile phone.

Step 1. Open CTS and ODK on the tablet
Step 2. Get participant phone number in CTS and enter it into phone
Step 3. Switch to ODK and open survey
Step 4. Make phone call
Step 5. Ask questions and enter answers into ODK
Step 6. At the end of the call, submit the survey and go back to Step 2

Volunteer participants were recruited, made live calls, collected their answers, updated their states on the CTS and ran the AVCS. The process ran as expected.

4.3.4 Installation and maintenance

Due to the system being online, there was no installation required. However, the project's Heroku account was upgraded from free dynos to hobby dynos to allow for the system to have an uptime of 100%. This would ensure that our dyno never went to sleep as this may have caused server downtime and reset. System click-through tests were run bi-weekly before surveys were to be conducted. Test surveys also were run to ensure the system was running as expected.

4.4 Data collection technology

Data collection was conducted on Google Nexus tablets. The same tablets were used by the enumerators to interact with the CTS and the AVCS. The tablets are Wi-Fi enabled and had no GSM. In order to make data collection relatively painless, we installed an open-source tool named Open Data Kit (ODK) on the tablets. ODK is a mobile device platform that allows for off-line, on-the-go surveys on android mobile devices [42]. ODK was chosen because it is free and open-source, it allows for off-line data collection, and the researcher was familiar with the tool.

ODK allows for a number of ways in which forms can be created. For this project we chose to use the Excel spreadsheet to design our surveys as shown in FIGURE 5. Once a form was created, it was converted in XML, which was then uploaded onto the ODK Aggregate server. Once the surveys were on the server, they would be available for use in the ODK application on the tablets. As shown in FIGURE 5, the server is also used as storage for completed surveys which are thereafter synched to the project's google drive where they would be until AVCS was ready to download them.

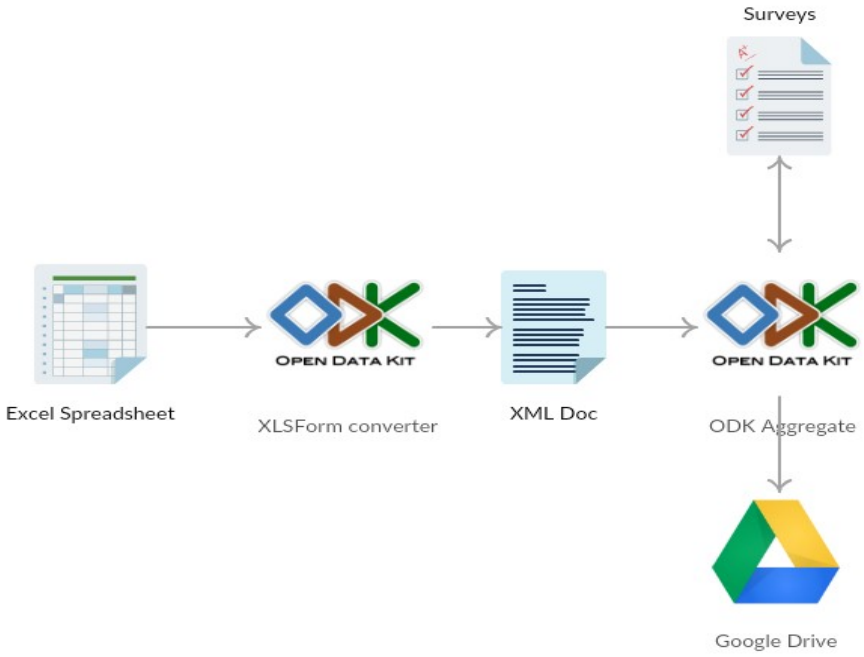

Figure 5: Survey generation to ODK

4.5 Call technology

Participants were called on a small android phone, the Samsung galaxy pocket. These the phones were chosen, as they were easily available and low-cost. Calls were made during surveys and during voice feedback calls. After hours, enumerators had to make calls from home and therefore they also used the phones as a Wi-Fi hotspot in order to interact with the Call Tracking System (CTS).

For the automated feedback calls, we used the Automated voice call system (AVCS). The AVCS was built using Twilio's API, which allowed us to programmatically make automated calls that delivered recordings to participants.

4.6 Text message delivery technology

There were two scenarios in which enumerators would send text messages to participants. The first was at the end of a survey call sprint. If a participant had not answered any of the enumerators call attempts, the enumerator would text the participant. The text was to inform the participant that there would be further attempts to contact them in the next survey round. The second scenario was in the feedback week. The airtime group would receive text messages from the enumerators thanking them for their participation and informing them of the airtime they had just received.

The text messages for workshop invitations were sent using a BulkSMS service. This allowed us to upload contacts in a CSV file and send a text message to all the contacts.

4.7 Mobile credit delivery technology

In order to send mobile credit to the airtime group participants, we used a service offered on the prepaid24 platform. It sends credit to multiple participants based on the information specified in a CSV. The CSV contains the participant phone number, their mobile network, and the amount to be credited.

5 Recruitment process and lessons learnt

Door-to-door recruitment was the data collection method of choice because it was ideal for our study requirements. The aim was to recruit participants to take surveys twice monthly. The following was required: (1) enough time to explain what the study was about, (2) ensure the household was located in the desired recruitment area, (3) establish a relationship during recruitment and (4) familiarise with locations through observation in order to inform the survey partners. In this chapter the recruitment steps and the lessons learnt during recruitment will be discussed.

5.1 Recruitment steps

During recruitment these steps were followed:

Step 1. Meet with enumerators and provide consent sheets
Step 2. Begin recruitment door to door
Step 3. Meet for an hour break and discuss current progress
Step 4. Continue with door to door recruitment
Step 5. Debrief session

Initially, enumerators recruited in pairs but were later split in order to cover more households in a shorter space of time.

5.1.1 Lessons learnt from recruitment day 1

On recruitment day 1, the enumerators had a final test run by recruiting one of their own households first. This proved to be successful as there was a boost in confidence after having recruited the first household. This household was not included in the data set due to the bias it would introduce.

It was surprising to discover that recruitment of 50 households took longer than the expected 4 hours. To solve this problem, the enumerators were split further so as to cover four households at

a time instead of two. To ensure the enumerators were still safe, a researcher or ACDI intern went out with them and played the role of “bodyguard”. The length of recruitment could also be attributed to the anger that was evident in the community towards the municipality.

A large proportion of the population left at home during recruitment hours was the elderly. This led to a few recruitment issues, mainly, some of the elderly were living with their children that worked during these hours. They claimed not to know enough about the household to participate and most of the elderly that lived in their own homes did not have mobile phones.

Some of the elder members of the community that we approached to participate did not seem to fully understand the aims of the project. The enumerators did, however, claim that providing example questions really helped people understand the project. Recruitment times were also an issue as most of the breadwinners in the household took the household’s only mobile phones to work.

People were very excited about the mobile credit incentive and this was highlighted by a story of one participant that claimed to be uncomfortable with giving out her number until she was told there would be a mobile credit incentive, to which she readily gave her mobile phone number. The electronic signature feature on Open Data Kit was also highlighted as a popular feature with participants.

On day 1 of recruitment, there was an estimated 30% success rate during recruitment. It was found that the recruitment success and failure rates needed to be recorded.

5.1.2 Lessons learnt from recruitment day 2

On day 2, the enumerators gained confidence when approaching strangers. A trend in effective recruitment times was also noted as shown in the table 4 below. It was observed that, people were particularly busy in the mornings and late afternoon. Prime recruitment time was in the late morning and early afternoon.

Table 4: Recruitment times and effects

Time	Enumerators	Participants
0900-1100	Cover a lot of ground	Unlikely to participate (chores)
1100-1500	Slowdown in recruitment speed	Likely to participate
1500-1700	Cover very little ground	Unlikely to participate (dinner preparations)

The success rate of day 2 was much higher than that of day 1 at 60%, which could be attributed to the FLOW supervisor joining on the recruitment trail. A phenomenon we came to describe as the "Loritha effect" was witnessed during this time. Loritha is a FLOW coordinator that manages to engage with people in a manner that seemed to lift people's emotions to a point where they wanted to participate in the project. The rapid pace of recruitment on day 2 made it appear that the team could potentially recruit the Control group and Voice group in one day.

5.1.3 Preparation for day 3 of recruitment

In order to prepare for recruitment on day 3 the routes on which the team would walk had to be mapped out so as to make our recruitment efficient. Everyone was split into two teams: the green team and the red team. Below is an example of how recruitment routes were mapped. This process made recruitment more efficient.

Figure 6: Example map of how recruitment paths were mapped. This is not the actual location where recruitment was done, but is used to illustrate the process of the two teams.

5.2 Day 3: Using lessons from day 1 and day 2

In order to achieve our goal, the day was scheduled to start very early; however, the team ran into an unforeseen issue. Half the desired participants were still shut in and asleep on our arrival. Once households had begun their days, with the "Loritha effect" still in full effect, recruitment of both participant groups was completed before sundown.

5.3 Recruitment feedback from enumerators

The enumerators identified 2 challenges in recruiting individuals: First, the enumerators indicated that the scripts they read from when introducing the project were too formal and had resulted in many participants not understanding the project. As a solution, in cases where people did not understand, the enumerators indicated the provision of examples helped participants grasp the purpose of project. Secondly, the enumerators highlighted that participants were worried about the

difficulty of the questions. As a solution, the enumerators explained that most of the questions were multiple choice and did not require them to explain.

We also took some time to discuss the differences in the towns. The enumerators felt that the Airtime group was fairly difficult during recruitment. This was particularly surprising as that was the treatment to receive mobile credit. This went against literature, as most literature talks of the positive effect monetary incentives have on participation [46, 49]. The enumerators credited this to the municipality that ran a survey in the community and had not taken steps to resolve problems result.

In closing, the enumerators mentioned that they felt safer and more confident when researchers were present during the recruitment. They felt that people were also easier to recruit when we were present as we were clearly community outsiders.

5.4 Recruitment numbers

In total, 200 households were recruited. As shown in table 5, the more the recruitment process was refined, the faster recruitment went. As shown in the table, recruitment hours went down and our success rate increased in the last two recruitment rounds. Starting by splitting the enumerators in order to cover more ground and going a step further on the final day by mapping the recruitment routes contributed to the speed. The increase in success rate could be credited to an increase in the confidence and experience of the enumerators.

Table 5: Recruitment outcomes

	Day	Participants	Success Rate	Recruitment (hours)
Airtime group	1	49	30% (Estimate)	8
Voice group	3	49	70%	3
Automated group	2	50	60%	6
Control group	3	52	67%	3

6 Evaluation Methodology

In order to answer the research questions, an in-depth look at each question was needed using statistical analysis. The first question—How do different incentives affect attrition? - was split into four sub questions that helped identify conclusive results. In answering the second question - How do different incentives relate to a participant's habits during a longitudinal survey? - six participant traits were identified that allowed comparison of how incentives influence participants. To answer question three - How did the different incentives affect the cost of the project? - qualitative feedback from workshops was gathered and a cost-benefit analysis was done.

6.1 Participation outcomes

The main purpose of the study was to monitor participant attrition. Specifically, we aimed to assess the effect of incentives on attrition. In order to differentiate between forms of attrition, the act of completely dropping out of the study was defined as *hard attrition*, and the act of skipping a survey week was defined as *soft attrition*. Non-participation was referred to as the act of not participating in the study in any way.

In order to evaluate the effects of incentives on participation, four questions were asked. The first two questions were:

1. How do the different incentives affect the likelihood of participants completing surveys?
2. How did the different incentive affect the likelihood of soft attrition?

To assess the likelihood of participants in either completing a survey or skipping a week of the survey, a categorical regression using a Cochran's Q test was run. Cochran's Q Test is a non-parametric repeated measures test that can be used with binary data. We chose Cochran's Q test because it allows a response variable to have one of two possible outcomes. In the case of this study we wanted to measure if any of our four incentives had an effect on the likelihood of a participant completing a survey or skipping a survey week. Cochran's Q test additionally allowed us to factor in the week as a variable as it allows for repeated measures.

The other two questions were:

1. How were the different incentives related to non-participation (NP)?
2. How did the different incentives affect hard attrition?

In order to assess non-participation and the effects of incentive on hard attrition, a categorical regression using a linear model was run. We chose to use this method because these are once off events, a participant cannot drop out of a study more than once. The model for this study was a first-order model with one binary predictor and is as follows: $y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$ where:

- y_i is the response of group member i. The responses are coded as follows:
    - 1 if a participant contributed/participated in the study.
    - 0 if a participant did not participate in the study
- x_{i1} is the group to which the member belongs.

6.2 Participation engagement

In order to categorize our participants' participation, some traits that we identified in our pool of participants was defined. Viewing the data in such a way would help give some context to the above analysis as well as fill in some gaps.

Participant behaviour was divided into six different groups: (1) active participants (AP) who answered four or more surveys without dropping out of the study, (2) active drop-outs (ADO), who completed four or more surveys but also dropped out of the study at some point, (3) wafflers (WAF), who did not drop out of the project but answered more than one survey and less than four surveys, (4) Low participation drop-outs (LPDO), who answered between one and four surveys and dropped out of the study, (5) drop-outs (DO), who asked to drop out at any point without answering any surveys, (6) no-contact (NC), who could not be reached during the course of the project.

Analysis was run in order to answer deeper questions about the effects of the treatments on participant engagement. The goal was to find which incentives attract a certain type of participant. The following questions were asked:

1. How were the different incentives related to active participation?
2. How were the different incentives related to active drop-outs?
3. How were the different incentives related to low participation drop-outs?
4. How were the different incentives related to waffling?
5. How were the different incentives related to drop-outs?
6. Was there any relation between no-contact and incentives?

To assess the participant engagement, a binary logistic regression was run using a generalized linear model. The model was a first-order model with one binary predictor and is as follows: $y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$ where:

- y_i is the response of group member i. The responses are coded as follows:
  - 1 if a participant was a member of the group, e.g. AP.
  - 0 if a participant was not a member of the group
- x_{i1} is the group to which the member belongs.

6.3 Project workshops

In order to get better insights into the mind-set and experience of the community, as well as to satisfy ethical obligations, workshops were held in each town in which surveys were conducted. The goal of the workshops was to ensure that all treatment groups were provided with the same feedback, monetary compensation and explanation for why this study was conducted. All the communities were Afrikaans speaking and therefore one of the FLOW coordinators was asked to help run the workshops. A systematic method to prepare for workshops was created: first, the researcher designed community specific slides for the workshop presentation. The slides and feedback data were reviewed in a dummy presentation. Next, the slides were sent to the FLOW coordinator. She would then review them in order to familiarize herself with the content. Lastly, all the stake-holders reviewed the slides with the coordinator on the day of the presentation.

Invitations were sent in a three-step process. First an invitation was sent by SMS using a bulkSMS service. The first SMS was sent five days before the workshop was to be held. Two reminder SMSes were sent eight hours and an hour before the actual workshop was to start.

The workshops were all recorded on a mobile device and started with the researcher welcoming the participants and thanking them for their participation in the project. The researcher then gave a brief overview of the whole project and why it was conducted. This was all in English and therefore it was kept brief. The FLOW coordinator would then take over and present the slides that were projected on a wall. After a section that usually consisted of four to five slides, there was a slide with questions. The questions and answers were both in Afrikaans. At the end of all slides, the floor was opened for participants to ask questions or give any feedback they had.

The day after the workshop, two things would happen. First, the coordinator would send an email with all the questions and answers from the workshop in English and second, the presentation was sent to the ward council. In the case of the three treatment groups that did not receive mobile credit, the mobile credit was sent based on how many surveys a participant answered.

6.4 Cost-benefit analysis

A cost-benefit analysis was conducted in order to understand not only the cost and benefits of the incentives, but also the social costs and benefits. The basic notion of a cost benefit analysis is to attempt to find out if a decision outweighs the benefits of the next best alternative. We therefore only make decision A if the benefits of decision A outweigh their costs and not otherwise [47]. In order to evaluate the cost and benefits of incentives, social costs and social benefits [30] needed to be calculated. Social costs for the purposes of this study involve: labour costs, operational costs and negative externalities. Social benefits were assessed for the participants, researchers, and the municipalities. Once costs and benefits were found, net present value or benefit/cost ratio was computed [30].

7 Results

This chapter examines the results of the methods discussed in the previous chapter, starting with an examination of participation outcomes, which will involve looking at how incentives affected participation. Then, how incentives affected participants will be explored. As a result, the likelihood of continued participation or dropping out based on incentive is assessed. These points are then tied into how incentives affected participation. We present results from the workshops that gave us insight into how participants felt about participation. Lastly, the cost benefit of incentives is analysed.

7.1 Participation Outcomes

At a high level the study results show that monetary incentives have an impact on performance, adding to prior work that has also found this to be true [11]. It is also seen that the monetary incentive resulted in lower non-participation and hard attrition in comparison to the other incentives. The voice incentive did not appear to do as well as the automated incentive in terms of completed surveys, but inversely, the automated incentive had a higher hard attrition rate when compared to that of the voice incentive as can be seen in table 6. The statistics presented in table 6 are further broken down and analysed in the next chapters using the tests described in chapter 6.

Table 6: Summary of participation outcomes

	Airtime	Voice	Automated	Control
Completed surveys per person	2.28	1.69	2.1	1.56
% Non-participation	34%	37%	34%	42%
% Hard Attrition	26%	35%	40%	50%

7.1.1 How do different incentives affect the likelihood of participants completing surveys

The results suggest that the first round of data is not influenced by incentives. This is due to the control group having the highest number of completed surveys. This was unexpected as it was hypothesised that the airtime group to have the most completed surveys every week. While this may be just noise in the data, it might be caused by two separate factors; (1) none of the groups had received any incentives yet, (2) participants that agreed to participate without incentive did not realise how burdening the surveys would be.

From week 2, however, a trend was observed that continued till the end of the project with the airtime group and the automated group topping the surveys and the control group having the least completed surveys in most weeks. The only contradiction to the hypothesis came in the form of the automated group that consistently completed more surveys than the voice group as can be seen in FIGURE 7. FIGURE 7 shows completed surveys as a percentage of participants in the group in a particular week.

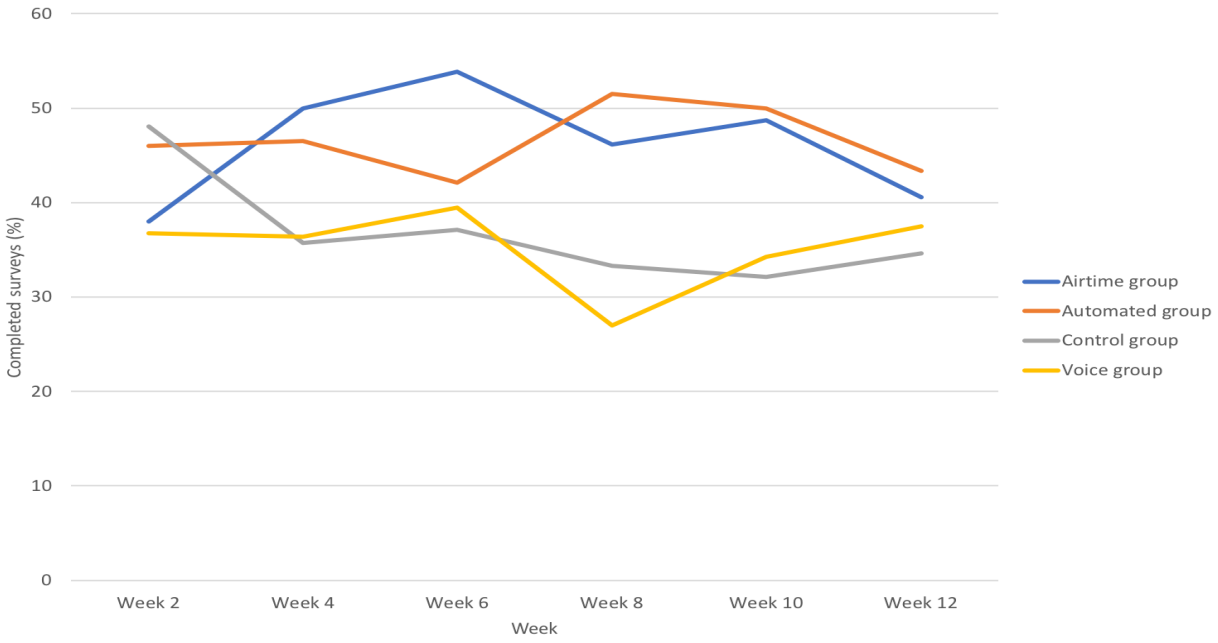

Figure 7: Number of Completed Surveys by week

In line with our hypothesis, it was found that the rate of survey completion decreased with time. Particularly interesting was the voice group's weekly non-participation reflecting in the number of completed surveys. We noticed that the high level of non-participation led to a decrease in the groups survey completion due to the almost directly proportional figures in the voice group. Noticeably though, there does not seem to be a direct correlation between non-participation and surveys completed as the automated group had the lowest non-participation rates for most of the project but was not the group with the highest survey completion.

Interestingly, because of the control group's high week one survey numbers, the control group's cumulative plot stays above that of the voice group up until week six as shown by FIGURE 8. This is interesting as the control group was expected to produce relatively low numbers in each survey week. This may suggest that given a longitudinal study with very few iterations, participants may not require an incentive in order to participate.

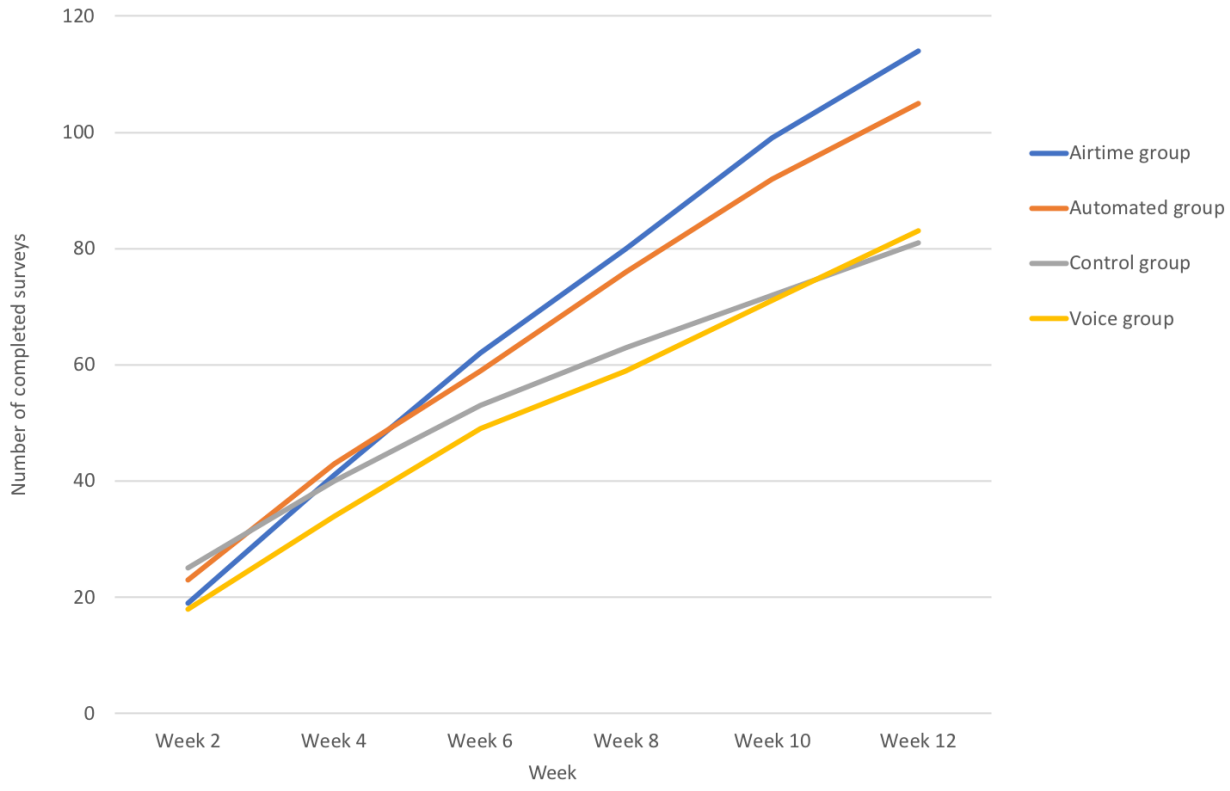

Figure 8: Number of Completed Surveys by week (Cumulative)

In order to find out whether there was a significant difference in completed surveys between treatment groups the Cochran's q-test was run and it was found that the control group was the only group that was significantly different as shown by its P-value that was less than 0.05:

Table 7:Cochran's q-test results for completed surveys

Treatment	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	P-Value
Airtime Group	0.333	0.438	0.417	0.375	0.375	0.292	NS
Voice Group	0.367	0.306	0.306	0.204	0.245	0.245	NS
Auto Group	0.451	0.392	0.333	0.333	0.333	0.275	NS
Control Group	0.442	0.288	0.250	0.192	0.173	0.154	<0.05

In table 7, we can also see that the probability of survey completion in the control group are significantly different to those of the other groups. These findings support the theory that incentives lead to an increase in performance [11].

7.1.2 How were the different incentives related to non-participation (NP)

In this study it was observed that non-participation ranged from 34% to 42%. As expected, the lowest number of non-participation was in the airtime group with the highest number in the out control group as shown in FIGURE 9. Non-participation is a major limitation to longitudinal studies as demonstrated in Binson's study that started with 5019 participants and got non-response from 1594 (32%) [10].

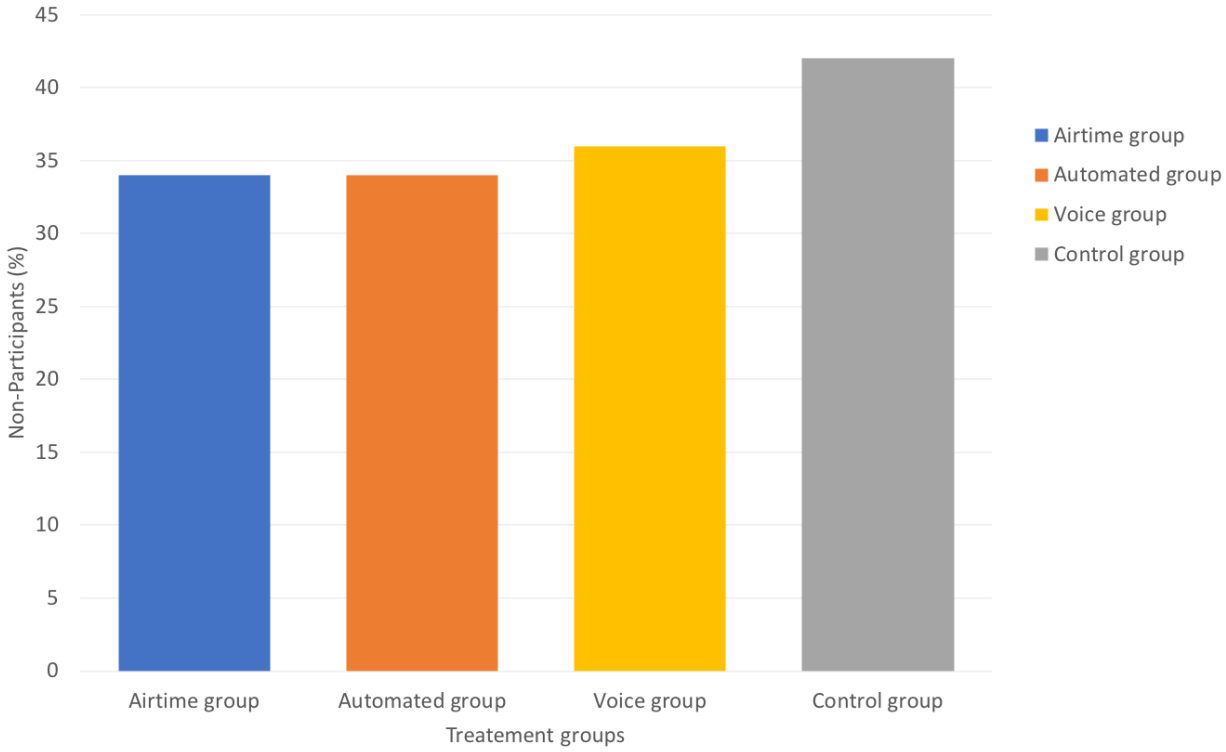

Figure 9: Percentage number of non-participants for entire duration of study

FIGURE 10 shows that non-participation was generally on the increase with time. The results also show that the voice group had the highest number of non-participants from week to week, however, overall the control group had the highest number of non-participants, that is, participants that did not answer a single survey.

The relatively high numbers of non-participants, per week, in the voice group to the high level could be due to soft attrition that was prevalent in the group. The participants were not participating in most of the survey rounds, choosing to skip certain weeks. They, however, were not dropping out at a rate equal to that of the control group. These high numbers of non-participants in the voice group can be seen in FIGURE 10 where we compare non-participation numbers per week.

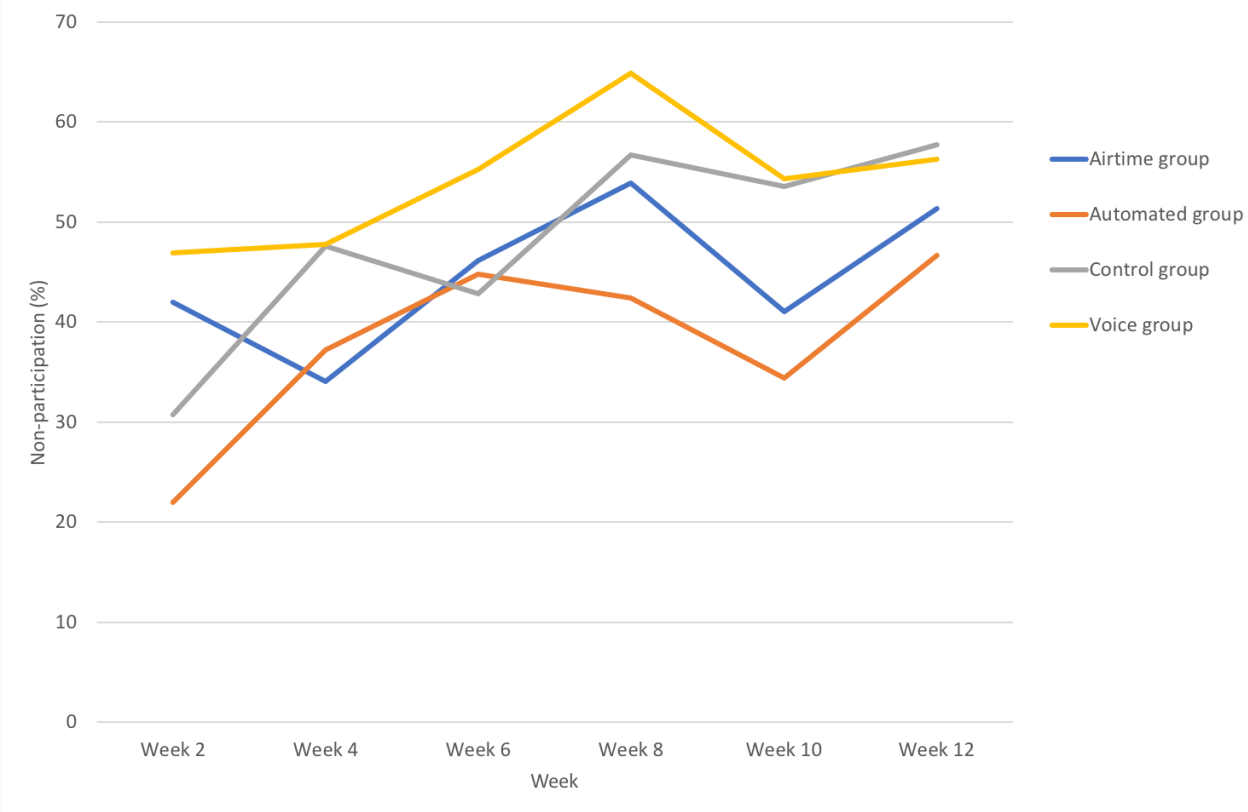

Figure 10: Percentage number of non-participants per week

To get more insight into whether incentives played a part in the differences observed, a generalized linear model was used and found no significance in incentives. This is attributed mainly to not having a significant number of participants in our study.

7.1.3 How did the different incentive affect the likelihood of soft attrition

The observations of Sauermann and Roach theorize that soft attrition shows some willingness to continue participating in a study but inability to answer the surveys on a particular day [64]. This study therefore assumes that participants that skipped a week instead of dropping out did so due to the relationship they had built with the enumerators and the incentive they were receiving.

As seen in FIGURE 11, it was observed that the automated group had the lowest percentage of participants that skipped a survey every week except week 2 and 3. The voice group had the highest number of participants that skipped a survey week. This, in part, could explain the difference in the two groups' number of completed surveys.

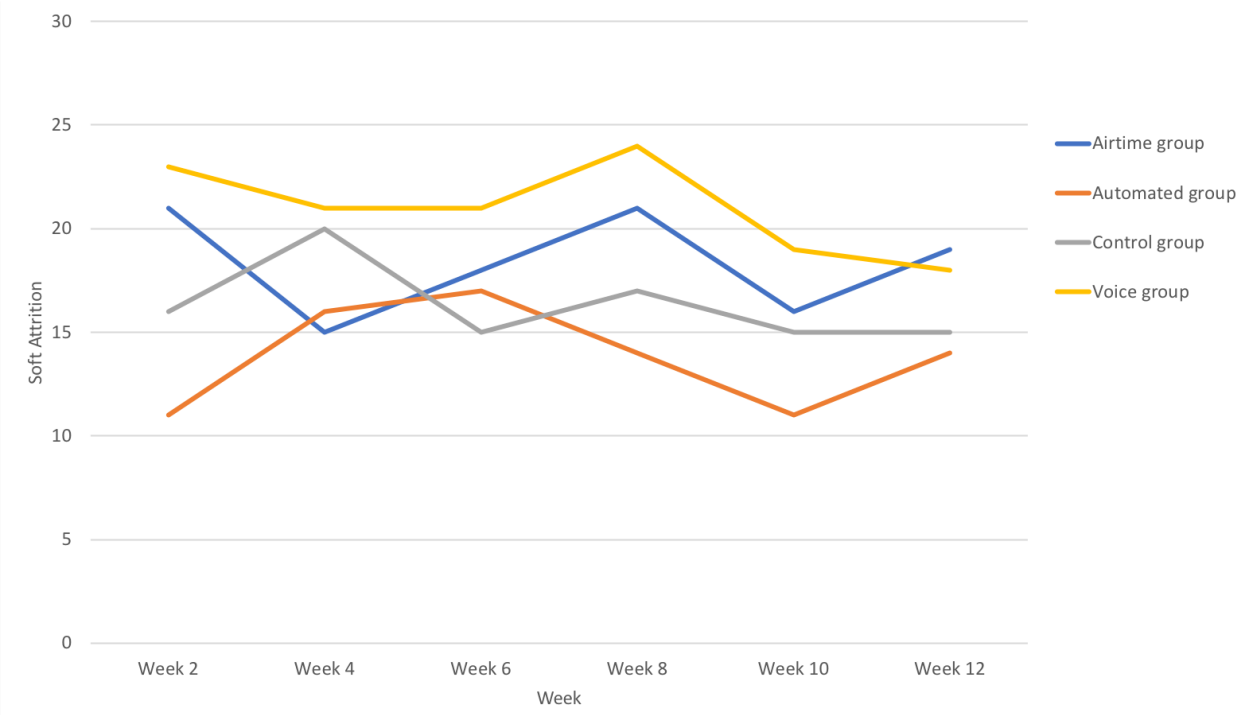

Figure 11: Number of people that were not contacted or asked to skip a survey by week

In the case of the automated group, it was observed that soft attrition by week was low but when calculated as a percentage of remaining participants, their soft attrition was high which can be seen in FIGURE 12. The inverse is true for the voice group as the results suggest that the voice group's participants were not dropping out of the study but asking to skip weeks instead when compared to the automated group.

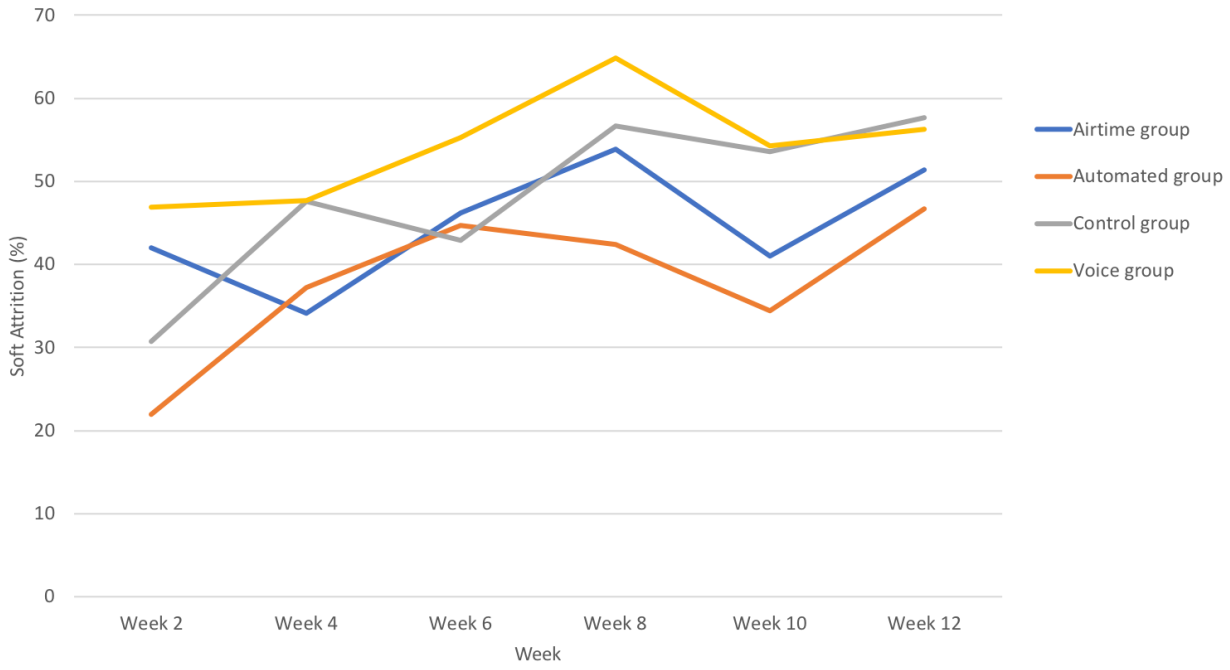

Figure 12: Soft attrition as a percentage of participants left in the project

In order to find out whether there was a significant difference in soft attrition between treatment groups the Cochran's q-test was run and it was found that the airtime group was close to significance when compared to the rest of the groups. It was also found that the probability of soft attrition was similar for all groups as shown in table 7.

Table 8: Cochran's q-test results for soft attrition

Treatment	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	P-Value
Airtime Group	0.542	0.332	0.354	0.396	0.354	0.375	<0.1
Voice Group	0.531	0.469	0.449	0.510	0.408	0.408	NS
Auto Group	0.412	0.373	0.333	0.314	0.275	0.275	NS
Control Group	0.365	0.385	0.327	0.346	0.327	0.327	NS

7.1.4 How did the different incentive affect hard attrition

Hard attrition was generally high in the first three survey weeks. Hard attrition rate was calculated as number of drop outs over the number of remaining participants. A decline in the rate of hard attrition was observed between week three and four. In the last three survey weeks, the hard attrition rates droped for all groups. This is supported by the paper by Tyler-Smith which speaks about attrition being highest at the start of e-learning courses [72].

This study defined the attrition Tyler-Smith refers to as "hard attrition", and as the results in FIGURE 13 show, the study had lower attrition rates, which could be as a result of the option given to participants to simply skip a survey week (soft attrition) [72]. Binson also supports these claims. He found that all his experiment groups experienced over 32% initial drop out [10].

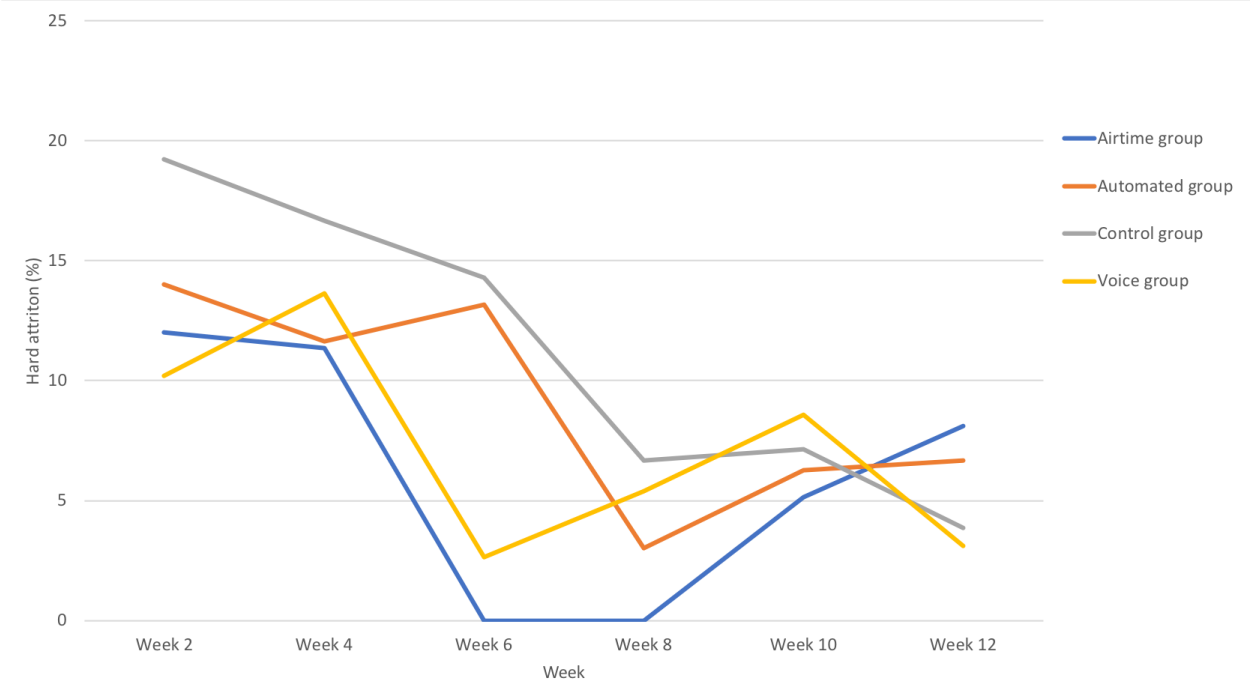

Figure 13:Hard attrition as a percentage of the number of people remaining in the project

Seeing as the recruitment of participants was in person, it is assumed that participants were unable reject the offer to participate in the project potentially due to the social stigma or social pressure, however, these participants would drop out with ease over the phone. As can be seen in FIGURE 14, there is a steep drop off in the graph during the first few weeks of the project which then levels out after week 3 and 4.

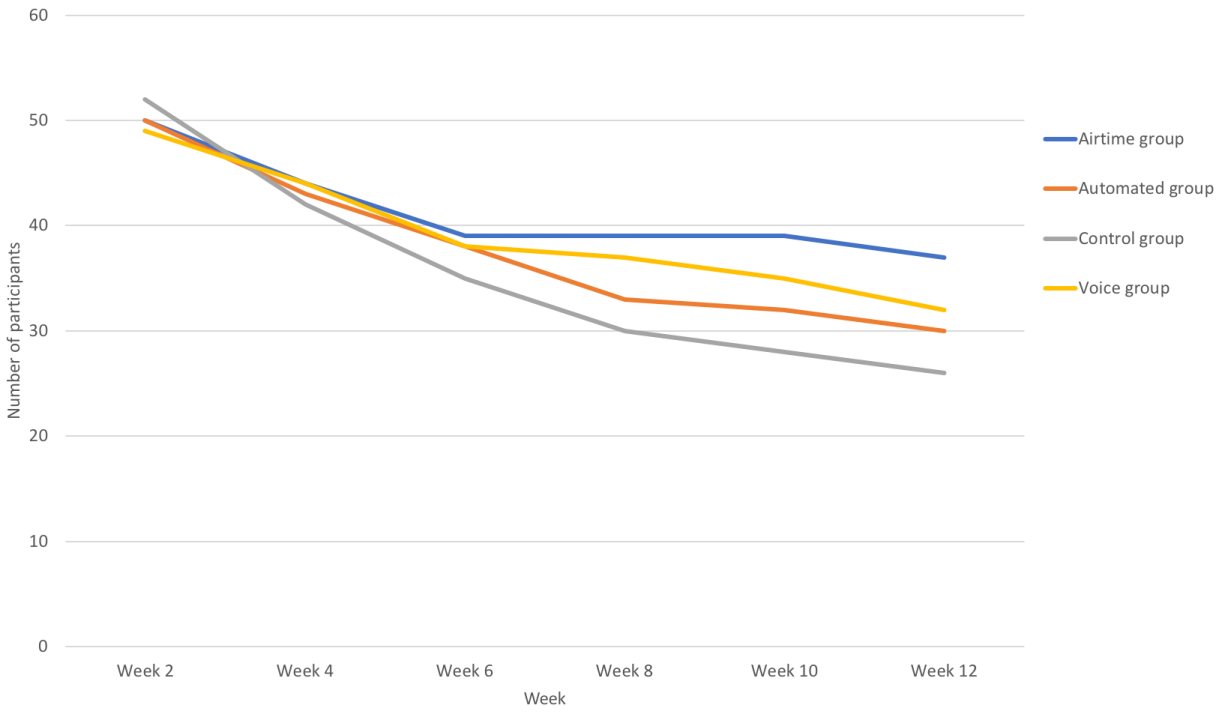

Figure 14:Participants left in the study due to hard attrition

In order to assess whether the incentives played a role in hard attrition, we used a generalized linear model and found no significance between the incentives, we however noticed trends with the airtime incentive ($Z=-1.863$, $p<0.1$) and voice incentive ($Z=-1.736$, $p<0.1$).

7.2 Participation engagement

As mentioned in previous chapters, it is important for longitudinal studies to have active participants in order to get a meaningful representation of the greater population. As shown in FIGURE 15, this study found that the airtime group had the most active participants, with the automated group displaying similar statistics. It was also observed that the airtime incentive did not have a larger group of participants in any other category. In order to assess whether incentive had an effect on participants becoming active participants a generalized linear model was used and found significance of incentive for the airtime incentive ($Z=2.059$, $p<0.05$). A trend for the automated incentive ($Z=1.685$, $p<0.1$) was observed and saw no significant difference between the voice and control. These results are supported by the hard and soft attrition results we observed above. Similarly, only the airtime group and the automated group had active drop outs, as can be seen in FIGURE 15. However, a generalized linear model determined no significance of incentive.

We defined wafflers as participants that answered more than one and less than four surveys and did not drop out of the study. As shown in FIGURE 15, the wafflers category (WAF) in the voice group and the control group displayed relatively high percentages compared to the airtime group and automated group. For the voice group, this showed a weakness in ability to entice participants to complete surveys but enough strength to keep them from dropping out. In addition, outside of the DO category, the control group's population was mostly found in the WAF category. For further insight a generalized linear model was used and found no significance of incentive.

Low participation drop-outs (LDPO) were participants that answered more than one survey and less than four surveys. In the LDPO category there was a trend with the airtime group having the lowest percentage falling under this category, with the control group having the highest. The groups that had the highest level of LPDO rates were the same groups that had a high level of hard attrition. Using a generalized linear model, no significance of incentive was found, however a strong trend for the airtime incentives ($Z=2.059$, $p<0.1$) was observed.

As seen in FIGURE 15, the DO category, once again a trend was seen from the airtime group to the control group. Relative to other categories, DO had very high numbers for all groups. Using a generalized linear model, no significance of incentive was found, however, given a larger sample group the data might have moved towards significance.

Lastly, the group no-contact category was assessed. The no-contact category was unlike the other groups as the incentive ideally has no effect on our ability to make first contact with participants. Using a generalized linear model, no significance of incentive was found as expected.

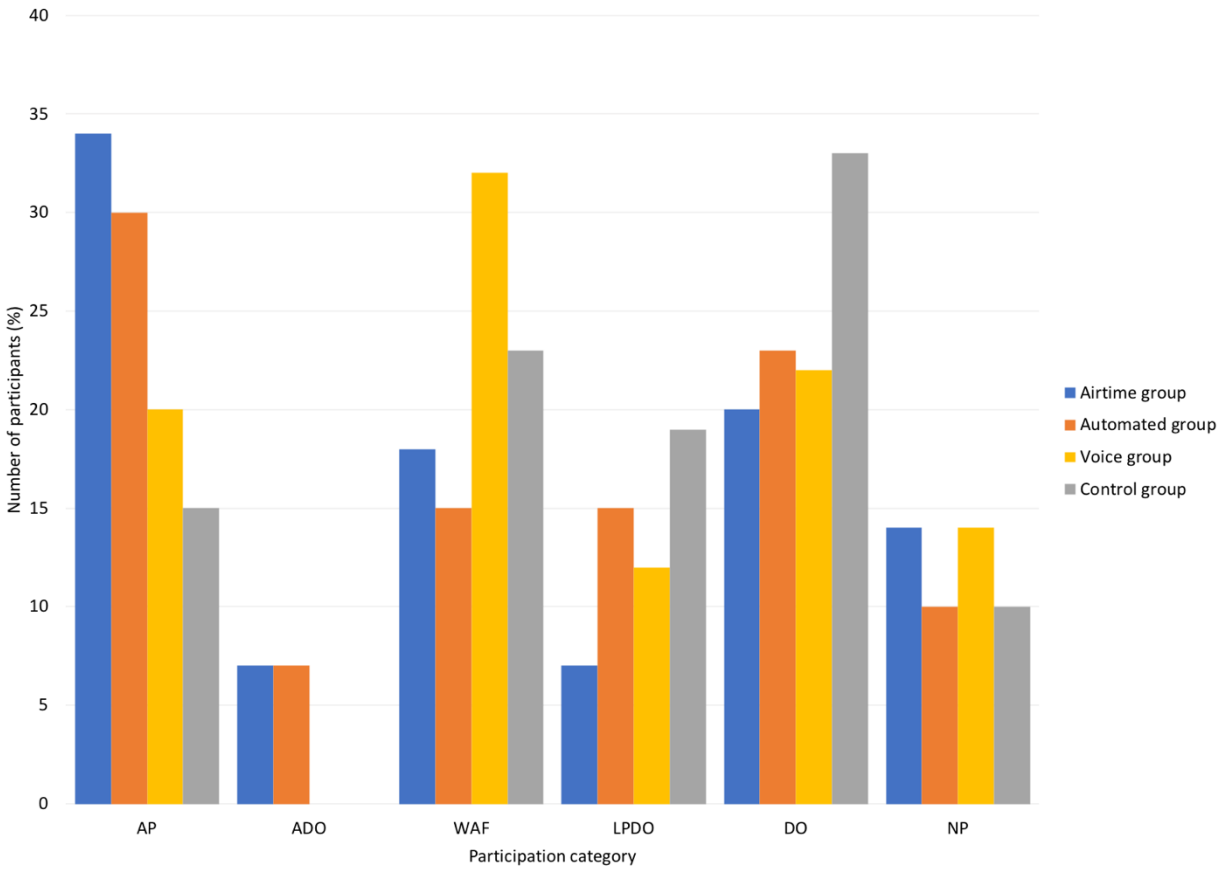

Figure 15:Participation Buckets

7.3 Project workshops

The workshops helped gather insight into how participants felt about the project. The anomaly of all the groups was the automated group that saw more ward committee members present than the participants. This happened as a result of attempting to join the ward committee meeting and our workshop into one meeting. In general, we found that the participants that received feedback were pleased with the consistent updates they were receiving. We noted that the airtime group was the least interested in the workshops, we however are unable identify why this was the case.

Table 9: Feedback from workshops

	Airtime group	Voice group	Automated group	Control group
Participants present	12	9	3	20
Committee present	0	1	10	4
Setting	Independent meeting	Independent meeting	Ward meeting	Independent meeting
Presentation time	30 Minutes	30 Minutes	20 Minutes	30 Minutes
Discussion time	10 Minutes	30 Minutes	0 Minutes	30 Minutes
Feedback	The airtime incentive motivated participation. Would participate again in similar studies.	First experience of feedback from a project. Happy about recieving feedback. Feedback workshop was of great value.	Stakeholder fatigue. Feedback was very useful to participants. Would participate again in a study that provides feedback.	Would have liked to receive regular feedback instead of at the end of the project. Study did not include issues of the youth.
Issues raised	Issue of data sensitivity were raised. Fear of giving sensitive information in the surveys.	Substance abuse was not covered extensively in surveys. It is the biggest issue in the community.	Skeptical about the real purpose project. Felt project is just a student project.	Skeptical about the real purpose project. Prefer face-to-face surveys.
Researcher comments	Little feedback from participants, workshop was not engaging enough.	Informal discussion before the workshop was helpful. Most engaged group in the after-presentation discussion.	Project participants did not show up. Distance between venue and community was a contributing factor.	Ward committee members wanted more people to participate. Participation motivated by possibility of change.

7.4 Cost-benefit of incentives

7.4.1 Project costs

The airtime incentives monetary costs were limited to the cost of paying the participants that completed surveys and the costs of the service we used to send airtime. The sending of airtime was done by the researcher and took a few hours to complete.

The voice incentives monetary costs included the cost of paying enumerators and buying the mobile credit. The enumerators were paid to make calls to participants providing feedback. They used the mobile credit to deliver the calls. The voice incentive also required training before every feedback session, which consumed time.

The automated incentive cost relatively little monetarily as the service we used was affordable. The only additional cost was the enumerator we paid to record the feedback. However, integrating the service with our system was not straight forward and as a result this incentive was expensive time wise.

The control group did not have any costs associated with it.

Table 10:Total costs of incentives

	Airtime	Voice	Automated	Control
Monetary Costs	*R*1300	*R*5150	*R*1000	*R*0
Time Costs	1 Day	2 Week	6 Weeks	0 Weeks

7.4.2 Project benefits

Airtime as an incentive is beneficial as it provided the highest number of completed surveys. Airtime provided the highest number of active participants. These two factors are particularly important to surveys as the richness of the data is dependent on the number of surveys completed. Additionally, the low cost and the little time required to pay participants is a significant benefit to

the researcher. The results also suggest that airtime works well as an incentive when conducting research with marginalized groups as it is of great benefit to them.

Voice on the other hand was not significantly different to the control, however it was observed that the participants in the voice group and the enumerators-built relationships over the survey iterations. The participants made this clear at the workshop with one participant asking for one enumerator personally. As a researcher, the interactive nature of the feedback is highly insightful as participants are able to respond to the information they are receiving.

The automated group had the benefit of delivering feedback to participants but the drawback of the feedback being non-interactive. The results however, suggest that this was enough to keep participants engaged as we saw a trend in active participants. There was also significant data in the completed surveys. The automated group was additionally very low cost to run.

The control group's benefits were limited to their low-cost nature. Table 11 shows the impact the incentives individually had on completed surveys and on active participants.

Table 11:Completed Surveys and Active Participants' significance

	Airtime	Voice	Automated
Completed Surveys	Significant	Not Significant	Significant
Active Participants	Significant	Not Significant	Trend

The social benefits were noted in the workshops. Both feedback groups gave positive reviews of the feedback process as they had been a part of other projects that never gave feedback. This proved to be of great benefit as participants were more aware of the water crisis that was affecting their area at the time and had new tips and tricks on how to save water as a result of the feedback they received. One participant said the direct feedback they had received allowed them to reflect on their communities, diets and usage of resources. The biggest social benefit was the participants' willingness to participate in future surveys as a result of the feedback they had recieved.

7.4.3 Analysis

From the costs and benefits outlined above, it can be advised that future researchers prioritize based on the resources they have and what they would like to get out of the project. If both time and monetary resources are present, it is advised to conduct a combination of automated and airtime incentives. If the researcher only has monetary resources available, then it would be advisable to proceed with an airtime incentive. In the presence of time resources and limited monetary resources, researchers should conduct automated feedback.

We are aware that there are cases that the researcher prioritizes information from participants, in that case, we would advise the researcher to conduct the voice incentive. However, the researcher must be weary of dropout rates and survey completion rates.

8 Discussion

8.1 Airtime is the most effective at minimising attrition

Airtime is the most effective incentive for minimising attrition and motivating participation. This point is backed by two facts, (1) the airtime group, displayed the least amount of attrition with an average hard attrition rate of 6.1% over the 12-week period of the surveys, (2) the airtime group had the most completed surveys with 114 in total.

Although airtime was the most effective incentive, feedback also had a positive effects on attrition. The voice group had a hard attrition rate of 7.3% whilst the automated group had an attrition rate of 9.1%. The control group had the highest hard attrition rate with a rate of 11.3%. The automated group, however, motivated participants to complete more surveys than the voice group despite the higher hard attrition rate. The automated group completed 105 surveys, the voice group completed 83 and the control group completed 81 surveys.

Whilst these results show us a clear preference for monetary incentives, they also suggest that participants preferred listening to feedback via a recording instead of a human. This may have been as a result of participants feeling that some of the information required was sensitive and by extension, the feedback was sensitive too.

Soft attrition, unlike hard attrition, did not show a visibly clear pattern. The airtime group had an average soft attrition rate of 44.7%, the automated group – 38%, the voice group – 54.2% and the control had 42,2%. The automated groups low rates of soft attrition explain why the group had higher levels of completed surveys. Participants in the automated group would either participate or drop out of the study leaving few to skip a survey week. We also found that soft attrition was highest in the voice group which highlighted that the participants receiving voice calls were motivated enough not to want to drop out of the project completely but not motivated enough to consistently complete surveys.

It was also found that a number of participants would pick up their calls, ask to be called later and subsequently switch off their phones. We realised this after a number of the enumerators reported this incident on several occasions. The frequency of this incident suggests it was not accidental or a simple case of phone batteries running out of power. Binson confirms, in his paper, that this is not a new finding and is common among survey practitioners [10]. These occurrences were marked as miscellaneous and also showed a pattern in motivation to participate. The project saw that this event happened three times in the airtime group, four times in the voice and five times in both the automated and the control group.

Based on the significant difference from the control group, airtime and feedback both have a positive effect on attrition. Airtime however, has a greater effect than feedback when trying to incentivise participation and minimise attrition in a longitudinal survey.

8.2 Airtime and automated feedback incite similar participant behaviour

The analysis of participant engagement revealed that regardless of high drop-out rates, the automated group had a high number of active participants similar to the numbers observed in the airtime group. Participants from the automated group either favoured dropping out completely or actively participating in the project. They had low levels of soft attrition which explains why the automated groups managed to have a high average of surveys per person relative to the other groups.

The voice group emerged as the top wafflers. This was unsurprising due to our expectation that the feedback incentive would motivate people not to drop out of the study completely. The voice group had very high levels of soft attrition and therefore a large percentage of their participants were wafflers. Also, unsurprising was the control group having the highest amount of drop outs due to the lack of incentives. This analysis also highlighted the problems associated with reaching participants. An average of 36% of participants could not be reached every week whilst 6% of our participant pool were never unreachable for the duration of the project.

It is evident that airtime and feedback have a positive effect on participant habits when compared to the control. Airtime and automated feedback resulted in active participants while voice feedback resulted in participants that did not participate as much but did not drop out of the study. We can also see that incentive has no effect on non-participation.

8.3 Automated feedback was the least expensive incentive and had statistical and social benefits

As shown in our analysis, automated feedback was the least expensive incentive. The most expensive incentive was the enumerator delivered feedback. This was almost five times more expensive than the airtime incentive which was the second most expensive incentive. Relative to the three other incentives, voice feedback had a negative effect on project costs as it raised project costs significantly.

The cost benefit analysis showed why researchers favour monetary incentives. They had a sufficiently low cost, little time spent setting up payments, a significant number of active participates and significant completed surveys. It is however important to note that the automated group was cheaper than the airtime group, was more time consuming to develop, had a trend in active participants, had a significant number of completed surveys and delivered feedback. The workshops gave us some insight into the significance of feedback to the participants. The participants saw the value of this information and while that cannot be measured or given significance, it was important in bridging the gap between researchers and community.

It was also evident that voice feedback might not be sustainable for larger longitudinal surveys when compared to airtime and automated feedback. The number of enumerators that would be required to deliver feedback as well as the amount of telephone credits that would be required to deliver all the calls would result in high project costs.

8.4 Implications of study on research

This study was important in showing researchers that with the use of technology, the act of giving feedback can be achieved at a minimal cost. Due to the use of technology, it was important not only to look at this study from a computer science perspective, but to inform researchers of when and how to utilise certain technologies in order to minimise cost and better the communities we work in.

The surveys in this study required use of ODK, with the integration of Google's APIs to streamline the process of getting data onto our system. The process of giving feedback required Twilio to be programmatically integrated into our system in order to eliminate the need for manual uploading of mobile numbers to Twilio. Both these processes made it important that this study be a computer science project as the feasibility of building such a system required assessment.

8.5 Future work

To continue this work, future studies would need to dive deeper into how modes of delivering surveys and incentives work together. There is a need to analyse whether certain incentives work with a specific mode of delivering surveys which would inform researcher when they attempt a specific type of survey. Is a face-to-face survey with a monetary incentive better than an over-the-phone survey with a feedback incentive? We believe these questions need to be answered if we are to build a framework that helps researchers choose the mode of delivery and appropriate incentive. While there is never a perfect solution, further research in this area would help guide researchers with make informed decisions.

9 Conclusion

In this dissertation we presented a study in which four sample groups, each receiving a different participation incentive, were compared in order to examine the effects of the incentive on attrition. This was done by conducting longitudinal surveys with the four groups. The study aimed not only to assess the effects of incentives on attrition but also to see how incentives affect the cost of a project and how incentives may potentially shape participant behaviour.

This paper showed the following:

1. Utilized a full longitudinal study design that was effective in an under-served environment (chapter 3).
2. Showed researchers how to build an efficient system for providing feedback without too many resources (chapter 4).
3. Shared lessons learnt during recruitment in under-served communities as well as feedback from enumerators that did the ground work (chapter 5).
4. Showed that feedback is a viable incentive given that time is available to develop a feedback system that works (chapter 7).

While it was easy to see why researchers tend to prefer monetary incentives as compensation—due to the effectiveness in reducing attrition—it is important to note that feedback can be used as an incentive with a lower capital investment, as was shown in the case of automated feedback. This study also showed that with a little time and no added cost, one can provide automated feedback to participants, while maintaining a similar attrition rate to that of monetary incentives.

Monetary incentives bring the ability to keep participants engaged while feedback bridges the gap between researcher and community. As shown by the airtime group in our project, monetary incentives are effective in keeping participants motivated, however as shown by our feedback

workshop, the feedback was invaluable to the communities in which we were working. We embarked on a project that would determine whether feedback was a stronger or comparable incentive to money. We found that feedback is comparable and has some advantages that money does not provide such as information exchange. As a result of this study, we recommend that longitudinal survey practitioners consider giving both monetary incentives and feedback due to the positives both provide. Based on our work with the communities, we believe strongly that researchers have an ethical and societal obligation to ensure that they are sharing the outcomes of their research, in a digestible manner, with the communities they study

Bibliography

[1] I. Alam, S. Khusro, A. Rauf, and Q. Zaman. Conducting surveys and data collection: From traditional to mobile and sms-based surveys. Pakistan Journal of Statistics and Operation Research, 10(2), 2014.

[2] H. Alderman, J. R. Behrman, H.-P. Kohler, J. A. Maluccio, and S. Watkins. Attrition in longitudinal household survey data: some tests for three developing-country samples. Demographic research, 5:79–124, 2001.

[3] R. M. Alvarez and C. VanBeselaere. Web-based surveys. The Encyclopedia of Measurement (forthcoming), 2003.

[4] C. S. Aneshensel, R. R. Frerichs, V. A. Clark, and P. A. Yokopenic. Telephone versus in-person surveys of community health status. American Journal of Public Health, 72(9):1017–1021, 1982.

[5] D. Ariely, A. Bracha, and S. Meier. Doing good or doing well? image motivation and monetary incentives in behaving prosocially. 2007.

[6] J. S. Armstrong and T. S. Overton. Estimating nonresponse bias in mail surveys. Journal of marketing research, pages 396–402, 1977.

[7] H. M. Badland, G. M. Schofield, K. Witten, P. J. Schluter, S. Mavoa, R. A. Kearns, E. A. Hinckson, M. Oliver, H. Kaiwai, V. G. Jensen, et al. Understanding the relationship between activity and neighbourhoods (urban) study: research design and methodology. BMC Public Health, 9(1):224, 2009.

[8] J. E. Barlett, J. W. Kotrlik, and C. C. Higgins. Organizational research: Determining appropriate sample size in survey research. Information technology, learning, and performance journal, 19(1):43, 2001.

[9] S. L. Bel, D. Chavernac, G. Mapuvire, and G. Cornu. Frontlinesms as an early warning network for human-wildlife mitigation: Lessons learned from tests conducted in mozambique and zimbabwe. The Electronic Journal of Information Systems in Developing Countries, 60(1):1–13, 2014.

[10] D. Binson and J. A. Catania. Random selection in a national telephone survey: a comparison of the kish, next-birthday, and last-birthday methods. Journal of Official Statistics, 16(1):53, 2000.

[11] S. E. Bonner and G. B. Sprinkle. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. Accounting, Organizations and Society, 27(4-5):303–345, 2002.

[12] R. F. Boruch and R. W. Pearson. Assessing the quality of longitudinal surveys. Evaluation Review, 12(1):3–58, 1988.

[13] M. J. Burke, S. A. Sarpy, K. Smith-Crowe, S. Chan-Serafin, R. O. Salvador, and G. Islam. Relative effectiveness of worker safety and health training methods. American journal of public health, 96(2):315–324, 2006.

[14] L. Calderwood. Reducing non-response in longitudinal surveys by improving survey practice. PhD thesis, UCL (University College London), 2016.

[15] S. K. Card, W. K. English, and B. J. Burr. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a crt. Ergonomics, 21(8):601–613, 1978.

[16] R. Chambers. Relaxed and participatory appraisal: Notes on practical approaches and methods. 2011.

[17] D. T. Chlup and T. E. Collins. Breaking the ice: using ice-breakers and re-energizers with adult learners. Adult Learning, 21(3-4):34–39, 2010.

[18] A. H. Church. Estimating the effect of incentives on mail survey response rates: A metaanalysis. Public opinion quarterly, 57(1):62–79, 1993.

[19] M. P. Couper and M. B. Ofstedal. Keeping in contact with mobile sample members. Methodology of longitudinal surveys, pages 183–203, 2009.

[20] K. Cox, R. Elkan, and N. Moghaddam. Feedback of trial results to participants: A literature review and stakeholder survey. In PSYCHO-ONCOLOGY, volume 18, pages 318–318. JOHN WILEY & SONS LTD THE ATRIUM, SOUTHERN GATE, CHICHESTER PO19 8SQ, W SUSSEX, ENGLAND, 2009.

[21] A. Dabalen, A. Etang, J. Hoogeveen, E. Mushi, Y. Schipper, and J. von Engelhardt. Mobile phone panel surveys in developing countries. 2016.

[22] G. Demombynes, P. Gubbins, and A. Romeo. Challenges and opportunities of mobile phonebased data collection: evidence from south sudan. World Bank Policy Research Working Paper, (6321), 2013.

[23] E. Deutskens, K. De Ruyter, M. Wetzels, and P. Oosterveld. Response rate and response quality of internet-based surveys: An experimental study. Marketing letters, 15(1):21–36, 2004.

[24] D. A. Dillman, G. Phelps, R. Tortora, K. Swift, J. Kohrell, J. Berck, and B. L. Messer. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (ivr) and the internet. Social Science Research, 38(1):1–18, 2009.

[25] D. A. Dillman, M. D. Sinclair, and J. R. Clark. Effects of questionnaire length, respondentfriendly design, and a difficult question on response rates for occupant-addressed census mail surveys. Public opinion quarterly, 57(3):289–304, 1993.

[26] D. A. Dillman, J. D. Smyth, and L. M. Christian. Internet, phone, mail, and mixed-mode surveys: the tailored design method. John Wiley & Sons, 2014.

[27] E. M. Falaris. The effect of survey attrition in longitudinal surveys: evidence from peru, côte d'ivoire and vietnam. Journal of Development Economics, 70(1):133–157, 2003.

[28] D. B. Fassnacht, K. Ali, C. Silva, S. Goncalves, and P. P. Machado. Use of text messaging services to promote health behaviors in children. Journal of nutrition education and behavior, 47(1):75–80, 2015.

[29] A. Fink. How to ask survey questions, volume 4. Sage, 2002.

[30] M. Florio, S. Forte, C. Pancotti, E. Sirtori, and S. Vignetti. Exploring cost-benefit analysis of research, development and innovation infrastructures: an evaluation framework. arXiv preprint arXiv:1603.03654, 2016.

[31] S. Fricker, M. Galesic, R. Tourangeau, and T. Yan. An experimental comparison of web and telephone surveys. Public Opinion Quarterly, 69(3):370–392, 2005.

[32] L. Fumagalli, H. Laurie, and P. Lynn. Experiments with methods to reduce attrition in longitudinal surveys. Journal of the Royal Statistical Society: Series A (Statistics in Society), 176(2):499–519, 2013.

[33] K. Ganesh, S. Mohapatra, S. Anbuudayasankar, and P. Sivakumar. User acceptance test. In Enterprise Resource Planning, pages 123–127. Springer, 2014.

[34] A. Georges, D. Schuurman, and K. Vervoort. Factors affecting the attrition of test users during living lab field trials. Technology Innovation Management Review, pages 35–44, 2016.

[35] J. Gerken. Longitudinal Research in Human-Computer Interaction. PhD thesis, 2011.

[36] M. Giugale, J. Hoogeveen, K. Croke, A. Dabalen, and G. Demombynes. Collecting high frequency panel data in africa using mobile phone interviews. Measuring African Development: Past and Present, page 186, 2016.

[37] M. Gleerup, A. Larsen, S. Leth-Petersen, and M. Togeby. The effect of feedback by text message (sms) and email on household electricity consumption: experimental evidence. The Energy Journal, pages 113–132, 2010.

[38] R. Gouveia, E. Karapanos, and M. Hassenzahl. How do we engage with activity trackers?: a longitudinal study of habito. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pages 1305–1316. ACM, 2015.

[39] C. Grady. Payment of clinical research subjects. The Journal of Clinical Investigation, 115(7):1681–1687, 2005.

[40] A. Gupta, W. Thies, E. Cutrell, and R. Balakrishnan. mclerk: enabling mobile crowdsourcing in developing regions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 1843–1852. ACM, 2012.

[41] V. Harte and J. Stewart. The reasons for attrition: we (still) haven't got a clue. 2010.

[42] C. Hartung, A. Lerer, Y. Anokwa, C. Tseng, W. Brunette, and G. Borriello. Open data kit: tools to build information services for developing regions. In Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development, page 18. ACM, 2010.

[43] Z. Hill. Reducing attrition in panel studies in developing countries. International journal of epidemiology, 33(3):493–498, 2004.

[44] C. Hsiao. Panel data analysisadvantages and challenges. Test, 16(1):1–22, 2007.

[45] P. A. Jessup and A. J. Stahelski. The effects of a combined goal setting, feedback and incentive intervention on job performance in a manufacturing environment. Journal of Organizational Behavior Management, 19(3):5–26, 1999.

[46] H. Laurie and P. Lynn. The use of respondent incentives on longitudinal surveys. Methodology of longitudinal surveys, pages 205–233, 2009.

[47] P. R. G. Layard. Cost-benefit analysis. Cambridge University Press, 1994.

[48] S. K. Lwanga, S. Lemeshow, W. H. Organization, et al. Sample size determination in health studies: a practical manual. 1991.

[49] P. Lynn. Methodology of longitudinal surveys. John Wiley & Sons, 2009.

[50] P. Lynn, N. Buck, J. Burton, A. Jäckle, and H. Laurie. A review of methodological research pertinent to longitudinal survey design and data collection. Technical report, ISER Working Paper Series, 2005.

[51] J. Magruder and N. Nattrass. Exploring attrition bias: the case of the khayelitsha panel study (2000-2004). South African Journal of Economics, 74(4):769–781, 2006.

[52] M. N. Marshall. Sampling for qualitative research. Family practice, 13(6):522–526, 1996.

[53] T. Mduluza, N. Midzi, D. Duruza, and P. Ndebele. Study participants incentives, compensation and reimbursement in resource-constrained settings. BMC medical ethics, 14(1):S4, 2013.

[54] S. Menard. Longitudinal research, volume 76. Sage, 2002.

[55] S. Molyneux, S. Mulupi, L. Mbaabu, and V. Marsh. Benefits and payments for research participants: Experiences and views from a research centre on the kenyan coast. BMC medical ethics, 13(1):13, 2012.

[56] H. Müller, A. Sedley, and E. Ferrall-Nunge. Survey research in hci. In Ways of Knowing in HCI, pages 229–266. Springer, 2014.

[57] C. Nicoletti and N. Buck. Explaining interviewee contact and co-operation in the british and german household panels. Technical report, ISER Working Paper Series, 2004.

[58] H. S. Parnes. Longitudinal surveys: prospects and problems. Monthly Labor Review, pages 11–15, 1972.

[59] S. Pedersen, A. Grønhøj, and J. Thøgersen. Texting your way to healthier eating? effects of participating in a feedback intervention using text messaging on adolescents fruit and vegetable intake. Health education research, 31(2):171–184, 2016.

[60] S. Rolstad, J. Adler, and A. Rydén. Response burden and questionnaire length: is shorter better? a review and meta-analysis. Value in Health, 14(8):1101–1108, 2011.

[61] T. Rosenstock, C. Lamanna, B. DeRenzi, S. Chesterman, S. Kadiyala, M. Ngendo, K. Chibuye, N. Choga, and M. Van Wijk. Surveillance of climate-smart agriculture for nutrition (scan): Innovations for monitoring climate, agriculture and nutrition at scale. 2016.

[62] D. L. Sackett. Bias in analytic research. Journal of chronic diseases, 32(1-2):51–63, 1979.

[63] T. A. Sanner and J. I. Sæbø. Paying per diems for ict4d project participation: a sustainability challenge. Information Technologies & International Development, 10(2):pp–33, 2014.

[64] H. Sauermann and M. Roach. Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. Research Policy, 42(1):273– 286, 2013.

[65] K. B. Sheehan. E-mail survey response rates: A review. Journal of Computer-Mediated Communication, 6(2):0–0, 2001.

[66] E. Singer, J. Van Hoewyk, N. Gebler, and K. McGonagle. The effect of incentives on response rates in interviewer-mediated surveys. Journal of Official Statistics, 15(2):217, 1999.

[67] D. A. Smith. Rapid software prototyping. 1982.

[68] P. Smith, P. Lynn, and D. Elliot. Sample design for longitudinal surveys. Methodology of Longitudinal Surveys, pages 21–33, 2009.

[69] R. Stringer. Twilio Cookbook. Packt Publishing Ltd, 2013.

[70] A. Team. Audacity (r): Free audio editor and recorder [computer program], 2013.

[71] D. Thomas, F. Witoelar, E. Frankenberg, B. Sikoki, J. Strauss, C. Sumantri, and W. Suriastini. Cutting the costs of attrition: Results from the indonesia family life survey. Journal of Development Economics, 98(1):108–123, 2012.

[72] K. Tyler-Smith. Early attrition among first time elearners: A review of factors that contribute to drop-out, withdrawal and non-completion rates of adult learners undertaking elearning programmes. Journal of Online learning and Teaching, 2(2):73–85, 2006.

[73] A. Vashistha, N. Kumar, A. Mishra, and R. Anderson. Mobile video dissemination for community health. In Proceedings of the Eighth International Conference on Information and Communication Technologies and Development, page 20. ACM, 2016.

[74] D. Watson. Sample attrition between waves 1 and 5 in the european community household panel. European Sociological Review, 19(4):361–378, 2003.

[75] N. Watson and M. Wooden. Identifying factors affecting longitudinal survey response. Methodology of longitudinal surveys, 1:157–182, 2009.

[76] H. Winschiers-Theophilus, T. Zaman, and A. Yeo. Reducing white elephant ict4d projects: a community-researcher engagement. In Proceedings of the 7th International Conference on Communities and Technologies, pages 99–107. ACM, 2015.

[77] P. Wohlstetter, A. Datnow, and V. Park. Creating a system for data-driven decision-making: Applying the principal-agent framework. School Effectiveness and School Improvement, 19(3):239–259, 2008.

[78] K. B. Wright. Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. Journal of computer-mediated communication, 10(3): JCMC1034, 2005.

[79] Y. Wu, M.-H. Chen, and J. Offutt. Uml-based integration testing for component-based software. In International Conference on COTS-Based Software Systems, pages 251–260. Springer, 2003.

[80] S. M. A. Zaidi, A. B. Labrique, S. Khowaja, I. Lotia-Farrukh, J. Irani, N. Salahuddin, and A. J. Khan. Geographic variation in access to dog-bite care in pakistan and risk of dog-bite exposure in karachi: prospective surveillance using a low-cost mobile phone system. PLoS neglected tropical diseases, 7(12):e2574, 2013.

[81] J. Zimmerman, A. Tomasic, C. Garrod, D. Yoo, C. Hiruncharoenvate, R. Aziz, N. R. Thiruvengadam, Y. Huang, and A. Steinfeld. Field trial of tiramisu: crowd-sourcing bus arrival times to spur co-design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 1677–1686. ACM, 2011.

[82] Tu, Huahong, Adam Doupé, Ziming Zhao, and Gail-Joon Ahn. "Sok: Everyone hates robocalls: A survey of techniques against telephone spam." In Security and Privacy (SP), 2016 IEEE Symposium on, pp. 320-338. IEEE, 2016.

[83] Stats SA. Mid-year population estimates. Pretoria, South Africa: Stats SA. 2013.