MONASH UNIVERSITY

DOCTORAL THESIS

Three essays on the evaluation of a poverty graduation program

Author:
Vilas GOBIN

Supervisors:
Dr. Paulo SANTOS
Dr. Russell TOTH
Associate Professor Andreas LEIBBRANDT

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Department of Economics

August 2016

Copyright Notice

Declaration of Authorship

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes three unpublished publications. The core theme of the thesis is poverty graduation programs. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Economics under the supervision of Paulo Santos, Russell Toth and Andreas Leibbrandt.

My contribution to the work involved the following:

Nature and extent of candidate's contribution
Formulation of research question, Design of survey, design of experiments, data collection, data analysis, writing and copy editing (60%).

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature:  **Date:** 22/08/2016

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main supervisor signature: 

Date: 22/08/2016

MONASH UNIVERSITY

Abstract

Monash Business School

Department of Economics

Doctor of Philosophy

Three essays on the evaluation of a poverty graduation program

by Vilas GOBIN

This thesis comprises three papers that examine a multifaceted approach to poverty alleviation implemented in northern Kenya. The approach combines multiple interventions with the aim of promoting entrepreneurship among ultra-poor women, and emphasizes cash transfers, in addition to business skills training, business mentoring and savings.

The first paper takes advantage of the randomized allocation of beneficiaries of the program to one of three funding cycles to estimate its impact on the welfare of beneficiaries. In the short-to-medium run participation in the program is found to have a positive and significant effect on income, savings, asset accumulation and food security.

The second paper looks at the impact of the program on female empowerment, and compares survey measures of empowerment to a measure derived from an incentivized decision making experiment. A positive impact of participation in the program on empowerment is found when using an experimental indicator, but not when using traditional survey measures. The experimental indicator also better correlates with indicators of well-being that are associated with more empowered women and is seen as a better measure of empowerment than survey measures, in this context.

The final paper takes advantage of the exogenous assignment of ultra-poor women to business groups to examine the effect of team heterogeneity on an experimental measure of trust. Heterogeneity in many characteristics is not found to affect the level of trust and trustworthiness between business partners. However, differences in asset wealth, measured by livestock ownership, is associated with less trust.

Acknowledgements

This thesis would not have been possible without the financial support of the Monash Business School, and the help of a large number of people who throughout my entire candidature have been nothing but supportive, encouraging me each and every day.

First and foremost, I offer my sincerest gratitude to my primary supervisor Paulo Santos whose passion for research inspired me to pursue a PhD. Paulo has always gone above and beyond, and one could not ask for more in a supervisor. His immense knowledge, enthusiasm, encouragement and friendship have helped me throughout my candidature. I would also like to thank the other members of my committee, Russell Toth and Andreas Leibbrandt for the assistance they have provided at the various stages of my candidature. This thesis would not have been possible without their expertise and guidance.

This dissertation also benefitted from feedback and comments from numerous persons including Gaurav Datt, Hee-Seung Yang, Asadul Islam, Dean Karlan, Lata Gangadharan, Lisa Cameron, Klaus Abbink and Choon Wang. Thank you for taking the time to read my work.

I must also acknowledge the BOMA Project for whom I worked as the Director of Monitoring and Evaluation from 2013 to 2014, during which time the data used in this dissertation was collected. Thank you for the opportunity to work with your organisation and for allowing me to use your data for my research. All analyses, interpretations or conclusions based on these data are solely that of mine and my co-authors. The BOMA Project disclaims responsibility for any such analyses, interpretations or conclusions.

A very special thanks is also extended to my colleagues, friends and family for their support and encouragement which has helped to keep me motivated. Thank you for your patience, understanding and love. Finally, I am forever indebted to my wife and best friend, Christine. Thank you for all of the sacrifices you have made. I could not have completed this journey without you.

Contents

List of Figures

List of Tables

For my parents

Chapter 1

Introduction

1.1 Introduction

This thesis comprises three papers that examine a multifaceted approach to poverty alleviation, the Rural Entrepreneur Access Project (REAP), targeted at ultra-poor women in northern Kenya. Such approaches to poverty alleviation are designed to simultaneously address the overlapping set of constraints facing the ultra-poor. One influential approach, pioneered by BRAC, is the Challenging the Frontiers of Poverty Reduction – Targeting the Ultra-Poor (CFPR/TUP). This approach is structured as a poverty graduation program: during a limited period (two years), its participants benefit from a set of interventions (initial consumption support and an asset transfer, together with savings services, skills training, and regular follow-up visits) with the expectation that, at the end of that period, participants would be able to participate in microfinance (Goldberg & Salomon, 2011; Matin, Sulaiman, & Rabbani, 2008).

The poverty graduation approach has been implemented and evaluated in many different contexts and has been found to result in sustained improvements across multiple outcomes including income, consumption, productive asset ownership and food security (Bandiera et al., 2016; Banerjee, Duflo, et al., 2015). This thesis contributes to this body of evidence through a randomised evaluation of REAP, a variation of the CFPR/TUP graduation approach, implemented in arid and semi-arid northern Kenya, a region where more than 80% of the population is estimated to be living below the national poverty line (Kenya National Bureau of Statistics and Society for International Development, 2013). The Rural Entrepreneur Access Project comprises a baseline package of interventions, including a USD 100 cash transfer to a group of three women to set up a microenterprise, business skills training, and business mentoring, which are followed, six months later, by a USD 50 cash transfer to the business conditional on having an active enterprise, and a focus on the importance of savings (training and introduction to savings groups). This sequence of interventions is targeted at ultra-poor women

and is designed to enable them to gain the assets and skills necessary to graduate from poverty, a motivation that is similar to the one behind the CFPR/TUP (BRAC, 2013).

Each of the three papers in this thesis makes use of data collected from a sample of REAP beneficiaries who were identified as being eligible for participation in the program in November 2012. In November 2012, local selection committees across 14 locations in northern Kenya identified 1755 women as being eligible for REAP. As part of REAP, these eligible women were required to form firms (composed of three women) which were then targeted with the multiple interventions that comprise REAP. Due to lack of capacity to simultaneously enrol all participants, it was decided to split the eligible women into three groups to be successively enrolled over the next three funding cycles (March/April 2013, September/October 2013 or March/April 2014). Assignment to each cycle was done randomly, through a public lottery that took place in each of the locations from which participants had been recruited, with one-third of the women enrolled in each funding cycle. All eligible women were interviewed at baseline (November 2012) and at two follow-up surveys, conducted at six month intervals and timed to coincide with the beginning of each new funding cycle.

The first paper takes advantage of the sequential roll-out of the program and the randomized allocation to each cycle to estimate the short-to-medium run impacts of REAP. After one year, the program is found to have a positive and statistically significant impact on income (34%), savings (131%) and asset accumulation (both consumer durables (29%) and productive assets (12.5%), and food security (21.5%). The primary channel for impact is through the setup of new petty trade enterprises, with time use data showing a corresponding tightly estimated reduction in leisure time and household activity. However there is a weak impact on monetary measures of consumption, and expenditure; if anything there is a small decrease in these variables in the medium-run (one year), at least at the upper percentiles of the wealth distribution, following the introduction and promotion of new savings groups. It is possible that in the medium-run, new asset accumulation and savings activities are absorbing the income increase. The program is also found to be highly cost-effective, with the average increase in household income covering the cost of delivering the program in just over one year.

In the second paper the focus turns to the impact of REAP on female empowerment. Female empowerment is seen as a key pathway by which to achieve impact, and many poverty alleviation programs target women as their main beneficiaries. But female empowerment dimensions are often not rigorously measured partly due to the fact that empowerment is not directly observable (van den Bold, Quisumbing, & Gillespie, 2013). Studies instead rely on proxy measures such as self-reported participation in household decisions or control over assets, but such measures may not be suited for particular local contexts (Schatz & Williams, 2012;

Upadhyay & Karasek, 2012) or may suffer from measurement problems such as social desirability bias (Jejeebhoy, 2002). In light of these issues, an incentivised decision making task is designed to directly capture a measure of empowerment.

Participants who were randomly assigned to start benefitting from REAP in March/April 2013 or March/April 2014 along with their spouses, were invited to participate in the experiments which took place in June/July 2014. In the experiment, women (and their spouses) perform two tasks. Firstly, they make decisions on investments in a simple risk elicitation task based on Gneezy and Potters (1997). In a second task, they play a modified version of the first task in which the investment decision takes place in two stages. In the first stage, the first mover decides how much to invest in the risky lottery. The second mover then has the opportunity to change the decision. Because all participants first act as the first-mover and then as the second-mover in response to every possible first move by their spouse, an indicator of empowerment that reflects the woman's decision to not compromise from her own preferred investment can be defined. This indicator explicitly reflects traditional concepts of power (Dahl, 1957): the woman, in making a decision to change the husband's investment decision to one that her husband would not otherwise choose (but she would) is exercising her power over her husband.

The experimental measure of empowerment is compared with traditional survey measures of empowerment but little correlation is found between these two types of measures. However, when these measures are correlated with indicators of well-being that one would expect to be associated with more empowered women, the experimental measure is found to correlate just as well or better with these indicators than the survey measures, leading to the conclusion that the experimental measure does capture empowerment. Participation in REAP is found to lead to increases in empowerment when empowerment is measured using the experimental measure, but not when it is measured using the traditional survey measures, with the experimental measure 31.5% higher for those who joined REAP one year earlier.

The third paper focuses on the three-person enterprises set up as part of REAP. Although the original criteria for group formation by business mentors specified that business mentors were to place all eligible women into groups of three, in November 2012 eligible women in some locations were assigned to business groups by local staff whereas in others they were allowed to form their business groups from the set of eligible women in their location. This paper takes advantage of the exogenous formation of business groups to examine the effect of heterogeneity on one important determinant of team performance, trust. In the context of rural East African communities where livestock management is the primary livelihood, we find that diversity in asset wealth, measured by livestock ownership, is associated with less trust as measured by the amount sent in an

experimental trust game. We posit that this behaviour in the game is guided by societal norms that underlie the exclusion of the poor from many important social networks.

Chapter 2

Poverty graduation with cash transfers: a randomized evaluation

2.1 Introduction

Microenterprises are the source of employment for more than half of the labor force in developing countries (de Mel, McKenzie, & Woodruff, 2008; Gindling & Newhouse, 2014), and are seen as potential engines of economic development by raising income of owners, creating a demand for labor and raising wages, and increasing market competition to generate lower prices for consumers (Bruhn, 2011; World Bank, 2012). Despite these potential benefits, many policymakers are concerned that some of the world's poorest people, sometimes known as the ultra-poor, are prevented from establishing such businesses or from participating in many popular approaches aimed at stimulating microenterprise formation.

Until very recently microfinance was advocated as a way to overcome financial market imperfections that limited the capacity of the poor to invest in profitable projects (Jolis & Yunus, 2003). However substantial recent evidence, using randomized control trials, points to the limited impact of microfinance on poverty alleviation, particularly for those in the lower tail of the income distribution, suggesting that alleviating credit constraints alone is not sufficient to reduce poverty through microenterprises (Banerjee, Karlan, & Zinman, 2015; Karlan & Zinman, 2011). This has prompted a shift in attention to other possible constraints, particularly entrepreneurial skills, knowledge and human capital, but results of evaluations of such interventions have been similarly mixed (e.g. Bruhn, Karlan, & Schoar, 2013; Drexler, Fischer, & Schoar, 2014; Valdivia, 2015).

Concerns around limited access to the microenterprise sector among the ultra-poor, that reflect the apparent lack of success of these "one-constraint-at-a-time" approaches, suggested the need for interventions that provide the ultra-poor with a localized "big push" to graduate from poverty by simultaneously addressing

the overlapping set of constraints that they face. One influential approach, pioneered by BRAC, is the Challenging the Frontiers of Poverty Reduction - Targeting the Ultra-Poor (CFPR/TUP). This approach is structured as a poverty graduation program: during a limited period (two years), its participants benefit from a set of interventions (initial consumption support and an asset transfer, together with savings services, skills training, and regular follow-up visits) with the expectation that, at the end of that period, participants would be able to participate in microfinance (Goldberg & Salomon, 2011; Matin et al., 2008).

Several recent impact evaluation studies provide promising support for this approach across a diverse set of developing countries. For instance, a randomized evaluation of CFPR/TUP across 1409 communities in Bangladesh finds that the program enabled ultra-poor women to engage in microentrepreneurial activities resulting in a 38% increase in earnings, which persists up to two years after participants graduate from the program (Bandiera et al., 2016). In a recent multi-site randomized evaluation across six countries, Banerjee, Duflo, et al. (2015) find similar impacts to those reported for Bangladesh: consumption, productive assets, income and revenue are higher in the treatment group at the conclusion of the program and remain higher one year after graduation. However, these impacts are found to be weaker in two study sites (Honduras and Peru), naturally raising questions about the external validity of the results.

Concerns about external validity are also present in another study in Andhra Pradesh, India, where Bauchet, Morduch, and Ravi (2015) evaluate a similar intervention and find no net impact on consumption, income or asset accumulation. The authors suggest that this result reflects mistargeting of individuals with strong labor market opportunities who quickly selected out of the program, which suggests broader lessons around muted impacts of ultra-poor programs when the opportunity costs to participation are relatively high, and how directed asset transfers could misdirect economic activity. There is also evidence that the asset transfers were liquidated and used to pay down debt, another source of targeting risk.

This paper presents a randomized evaluation of the Rural Entrepreneur Access Project (REAP), a variation of the CFPR/TUP graduation approach, implemented in arid and semi-arid northern Kenya, a region where more than 80% of the population is estimated to be living below the national poverty line (Kenya National Bureau of Statistics and Society for International Development, 2013). REAP comprises a baseline package of interventions, including a USD 100 cash transfer to set up a microenterprise, business skills training, and business mentoring, which are followed, six months later, by a USD 50 cash transfer conditional on having an active enterprise, and a focus on the importance of savings (training and

introduction to savings groups). [1] This sequence of interventions is targeted at ultra-poor women and is designed to enable them to gain the assets and skills necessary to graduate from poverty, a motivation that is similar to the one behind the CFPR/TUP (BRAC, 2013).

This program, while similar in spirit to ultra-poor programs that have been implemented elsewhere, also has a number of notable differences. First, contrary to most other ultra-poor programs, the program relies entirely on the transfer of cash rather than of a physical asset as a way to increase beneficiaries' wealth. (e.g. Bandiera et al., 2016; Banerjee, Duflo, et al., 2015; Bauchet et al., 2015). Although cash transfers have the potential advantage of providing increased flexibility, by allowing beneficiaries to decide on the nature of their investment, they have played a minor role in these programs given concerns about their possible misuse (consumption, payment of existing debt). Cash, when delivered, is mostly conceptualized as consumption support, intended at preventing beneficiaries from "eating" their assets (sometimes literally, in the case of livestock transfers). This concern is potentially more important in the case of REAP given that there was no provision of initial consumption support. Our results show that the structure of the program (in particular, we suspect, the conditionality of the second grant and mentor input), seem enough to direct women toward an enterprise investment.

Second, the program is explicitly enterprise focused, with the requirement that women form three-person groups to jointly run the enterprise. This may provide additional social support and help the enterprises reach a viable scale, while providing additional accountability around the use of grant funds, but may also introduce additional costs in running a business that may, ultimately, be detrimental to its success. Finally, and not necessarily less important, REAP is implemented in the context of very limited market access, with an economy that is based on one activity (raising livestock) that is prone to frequent shocks due to drought. In this context, and in contrast to settings such as that studied by Bauchet et al. (2015), participants are unlikely to have even the prospect of other remunerative opportunities, suggesting lower risk of mistargeting of the intervention.

While the program differs from other ultra-poor programs in some respects, the findings are qualitatively similar. We find that, after one year, this program has a positive and statistically significant impact on income (34%), savings (131%) and asset accumulation (both consumer durables (29%) and productive assets (12.5%)), and food security (21.5%). The primary channel for impact is through the setup of new petty trade enterprises, with time use data showing a corresponding tightly estimated reduction in leisure time and household activity.

[1] The program is implemented through a NGO, The BOMA Project. See http://bomaproject.org/the-rural-entrepreneur-access-project/ for a complete description of REAP.

However we find a weak impact on monetary measures of consumption, and expenditure; if anything there is a small decrease in these variables in the medium-run (one year), at least at the upper percentiles of the wealth distribution, following the introduction and promotion of new savings groups. It is possible that in the medium-run, new asset accumulation and savings activities are absorbing the income increase. We also find the program to be highly cost-effective, with the average increase in household income covering the cost of delivering the program in just over one year.

Our results are similar to those presented in Blattman, Green, Jamison, Lehmann, and Annan (2016), in an analysis of a program that shares important similarities with REAP, as it also focuses on enterprise development through a cash transfer (USD 150), short business training, and ongoing supervision. They find that, over a similar evaluation horizon to ours, the program leads to an important increase of microenterprise ownership, mostly in petty trade, and income. They argue that, even with no consumption support and in a context of arguably little accountability around the use of grant funds, recipients are remarkably compliant in directing the funds to enterprise formation (rather than immediate consumption, as feared). A further treatment encouraging the formation of self-help groups, implemented in half of the treatment villages, led to a doubling of the reported earnings of those receiving the additional treatment, with most of the impact apparently due to increased informal finance and economic cooperation, a result that is suggestive of the additional importance of deeper financial services (insurance, in this case; savings, in the case of REAP) in buttressing such interventions.

The results presented in our article complement the recent evidence on ultra-poor interventions, while providing additional corroboration of external validity in a particularly remote and economically-challenging setting. In contrast to ultra-poor programs that focus on a relatively narrow set of enterprises, selected by the implementers of those programs, the REAP program provides a wider agency in how beneficiaries use the relatively small transfers that they receive. Despite this notable difference, the impact results match up relatively well with recent ultra-poor programs on outcomes, with notable increases in income and assets and little impact on consumption in the initial stages of the program. The pathway of livelihood change is also quite clear, as underemployed women shift away from leisure and household activity and into remunerative petty trade. This suggests some robustness in the implementation of such programs, with room for experimentation in program design in future iterations, for example in using group-based approaches, transferring cash rather than an asset (which can greatly reduce implementation costs), or reducing costs by minimizing initial consumption support.

The remainder of the paper proceeds as follows. In the next section we provide

a detailed description of REAP before presenting the identification strategy and the data used in this paper. We are able to take advantage of the randomized roll-out of the program, which resulted from over-recruitment during the participant selection stage of the program, to obtain unbiased estimates of the program's impact on household welfare. We next present results of tests of the assumptions underlying the identification strategy before discussing spillover and anticipation effects. This is followed by the presentation and discussion of the main results.

2.2 Overview of the intervention

The Rural Entrepreneur Access Project was implemented in 14 locations in the southern and central parts of Marsabit County, in the Arid and Semi-Arid Lands (ASALs) of northern Kenya (see Figure 2.1), a region where more than 80% of the population are estimated to live below the national poverty line (Kenya National Bureau of Statistics and Society for International Development, 2013). [2] The main livelihood option in these locations is pastoralism, with livestock serving both as a source of income and food for herders and their families. Pastoralism, however, is highly susceptible to weather and other shocks, and repeated droughts frequently have devastating impacts on households' livelihoods (Silvestri, Bryan, Ringler, Herrero, & Okoba, 2012), resulting in many households no longer being able to meet their basic needs due to the loss of herds from which it is hard to recover (Barrett & Santos, 2014; Lybbert, Barrett, Desta, & Coppock, 2004). Such households are forced into begging, unskilled wage labor, different forms of petty trade, and become reliant on food aid to meet their dietary needs. [3]

Opportunities to engage in non-pastoral activities are further restricted by the fact that communities in this region tend to be excluded from national development processes, have low population densities, have limited access to markets or other infrastructure, and face financial and human capital constraints (Elliot & Fowler, 2012). By targeting the poorest women in these communities, REAP aims to provide the most vulnerable households with a pathway out of poverty by alleviating the financial and human capital constraints that they face.

[2] In 2005/06, the poverty line was estimated at Ksh 1,562 (PPP USD 77.07 at 2014 prices) per adult equivalent per month for rural households (Kenya National Bureau of Statistics, 2007). In 2009 it was estimated that nationally, 45.2% of the population lived below the poverty line (Kenya National Bureau of Statistics and Society for International Development, 2013).

[3] Little, McPeak, Barrett, and Kristjanson (2008) examine different proxies for poverty and welfare in northern Kenya. They identify poverty as being most prevalent among sedentary households that are no longer directly involved in pastoral production or are in the process of exiting pastoralism. These households have little or no livestock and tend to be involved in unskilled wage labor and petty trade.

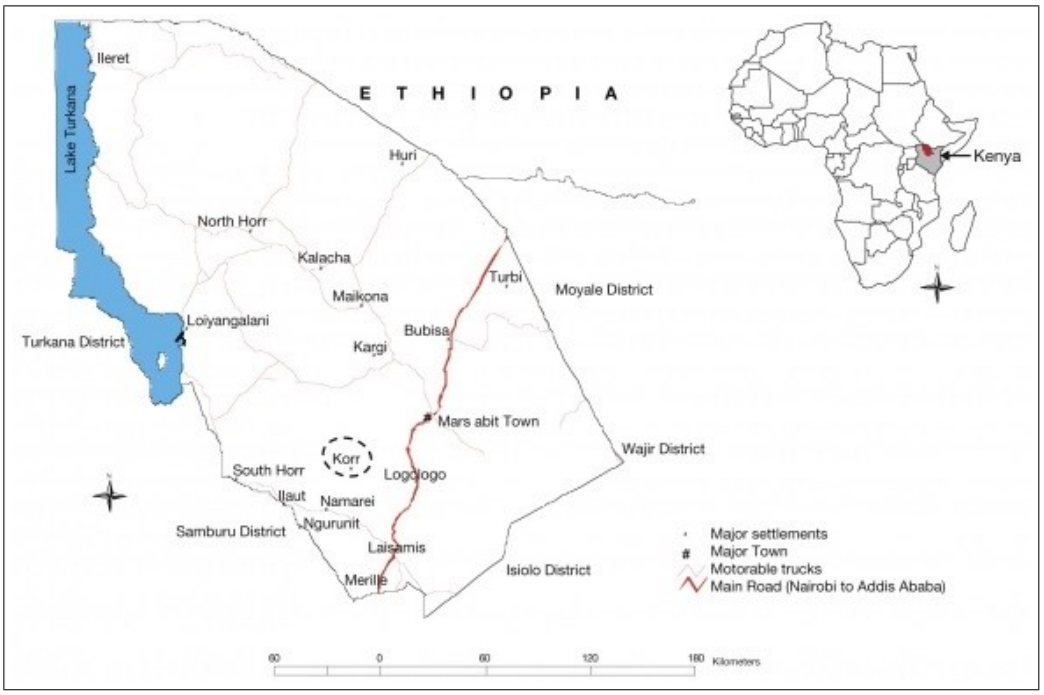


FIGURE 2.1: Map of Marsabit County (Warui & Kshatriya, 2009).

2.2.1 Structure and timing of the program

The main aim of REAP is to graduate ultra-poor women from poverty, through a set of interventions that include the development of business plans and mentoring, grants and access to saving mechanisms. The sequence of these interventions is presented in Figure 2.2, and each intervention is briefly described below.

Participant Selection. Program eligibility is determined by local committees, formed specifically for targeting. [4] These committees were asked to identify women who were among the poorest of the poor in the community, prioritizing those with no other sources of income besides the business to be formed, who were also considered to be responsible and entrepreneurially minded, and were willing to run a business with two other women. [5] Trained business mentors ensured that the local committees followed these criteria when selecting participants. [6] Once the participants were selected and accepted the invitation to participate in REAP, the business mentor proceeded to form business groups of three women.

[4] The committees generally comprise ten persons, with equal representation of clans and ethnic groups in the community, and with at least half of them being women.

[5] In addition, and recognizing the importance of inter-ethnic rivalries in northern Kenya, selection committees were asked to select participants in order to lead to an equal representation from various clans and ethnic groups and appropriate representation of persons from the town center and more distant villages. Finally, immediate relatives of any BOMA Project staff were considered ineligible. More recently, participant selection procedures included a Participatory Wealth Ranking to identify the poorest, followed by a short interview, used to confirm eligibility.

[6] Mentors are employed at the location level and, prior to the recruitment of participants, participated in a training of trainers program which lasted for five days. Each location comprises many sub-locations which are formed by smaller villages, known as manyattas.

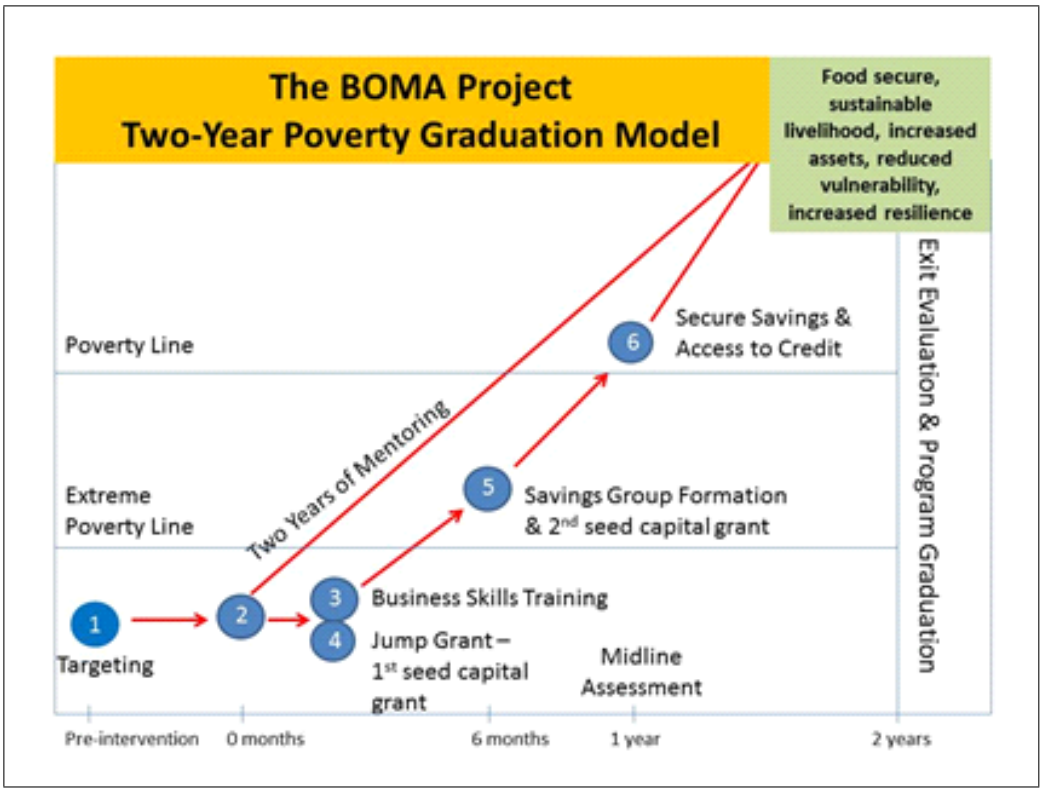


FIGURE 2.2: The six steps of REAP (The BOMA Project, 2014).

Business Planning and Business Skills Training. In the month leading up to program enrollment, the business mentors met with beneficiaries to assist with the development of a business proposal. The mentor was expected to get a better understanding of the group members' abilities and previous business experience before going through the basics of setting up a business with the group. On the day of program enrollment, all participants were required to attend a short business skills training session, delivered by mentors under the supervision of REAP field officers. [7] Over the course of the program participants benefited from approximately 17 hours of ongoing training. [8]

First Grant and Business Mentoring. At the end of the business skills training session business groups were provided with a cash grant of USD 100 (PPP USD 237.97 at 2014 prices) to be used to establish their business, an amount which is equivalent to approximately 7.5 months of expenditure per capita. [9] Once the

[7] These two sessions took approximately four hours to complete and covered the following content: accounting, financial planning, product ideas, marketing, pricing and costing, inventory management, customer service, business investment and growth strategies, employee management, savings, and debt.

[8] This included a half day training on savings that took place six months after the business training, and nine one-hour training sessions that took place during savings group meetings.

[9] From hereon in, all monetary values reported in the paper are in PPP terms at 2014 prices unless otherwise stated. We use the following PPP exchange rates to convert Kenya Shillings to USD PPP: 36.83 (2012), 38.38 (2013), 40.35 (2014). These values are then converted to 2014 prices by multiplying the ratio of the 2014 US Consumer Price Index (CPI) to the US CPI for the relevant year.

groups received their grants they were free to invest the money, including by making changes in their initial business proposal.

The distribution of the initial grants was followed by a period during which a mentor regularly met with the business group (at least once a month) to monitor its progress and offer advice and training. The role of the mentor was to help in the start-up of the business, through the provision of information (such as where to source goods and market conditions). Additionally, it was expected that, by providing ongoing training and support, the mentor would help the group with record keeping and, if needed, in managing conflicts within the group. Mentoring would last until groups formally exited the program, two years after its start, and over the course of the program each business was expected to benefit from approximately 30 hours of mentoring.

Second Grant, Savings Training and Savings Group Formation. Six months after the start of the business, groups were eligible for a follow up grant of USD 50 (PPP USD 118.98) conditional on meeting the following criteria: two or more original members remained involved in the business; members held business assets collectively; and the business value (defined as the sum of cash on hand, business savings and credit outstanding, and business stock and assets) was equal to or greater than the value of the initial grant. Participants were also required to participate in a short training session on savings, designed to provide a basic understanding of the formation and operation of savings groups including their rules, record-keeping, and issuing of loans. These conditions were known by participants since the start of the program.

After the savings training and the second grant distribution, participants were encouraged to form a savings group (SG) or join existing ones. The decision to join a group was both non-compulsory and individual (i.e., it was not a business group decision). The savings group model introduced to participants during the training most closely resembled Village Savings and Loans Associations (VSLA), also known as Accumulating Savings and Credit Associations (ASCAs), described in Allen (2006). The groups are self-managed and allow members to save money and access loans which are paid back with interest.

2.3 Research design

In this section we provide details of the random allocation of participants to treatment and control groups. We also report on tests of the assumptions underlying the identification strategy and discuss spillover and anticipation effects.

2.3.1 Randomization of program assignment

In November 2012, the local selection committees across 14 locations in northern Kenya identified 1755 women as being eligible for REAP. Due to lack of capacity to simultaneously enroll all participants, it was decided to split the eligible women into three groups to be successively enrolled over the next three funding cycles (March/April 2013, September/October 2013 or March/April 2014, hereafter groups *A*, *B* and *C*, respectively). [10] Assignment to each cycle was done randomly, through a public lottery that took place in each of the locations from which participants had been recruited, with one-third of the women enrolled in each funding cycle. [11] A public lottery was used to ensure that the allocation to funding cycle was transparent and fair, and seen as such. The random assignment of the beneficiaries to each cycle, if not defied, should lead to balanced groups. All eligible women were interviewed at baseline (November 2012) and at two follow-up surveys, conducted at six month intervals and timed to coincide with the beginning of each new funding cycle. [12]

None of the eligible participants declined to participate in the program, or was allowed to participate outside of the group to which they were randomly allocated. Survey attrition is very low in both follow-up rounds of survey. Less than 2% of women could not be reached for a follow-up interview in either the midline or endline surveys (see Table 2.1).

TABLE 2.1: Sample sizes (individuals and businesses)

	Group A		Group B		Group C	
	# Women	# Businesses	# Women	# Businesses	# Women	# Businesses
Baseline (Nov 2012)	585 (100%)	195 (100%)	585 (100%)	195 (100%)	582 (100%)	194 (100%)
Midline (Sep 2013)	549 (93.8%)	186 (95.4%)	565 (96.6%)	193 (99.0%)	565 (97.1%)	193 (99.5%)
Endline (Apr 2014)	534 (91.3%)	189 (96.9%)	556 (95%)	192 (98.5%)	561 (96.4%)	190 (97.9%)

Together, the sequential roll-out of the program, the randomized allocation to each cycle, the perfect compliance of observations to treatment and control groups, and the extremely low attrition rate, allow us to identify the program impacts in a relatively straightforward way.

[10] As a result, sample size was determined by the capacity of the program to reach participants. We conduct *ex post* power calculations to determine if there is sufficient power, given the predetermined sample size, to reliably estimate program impacts, and find that in most cases the minimum detectable effect size is as low as 15%.

[11] Initially 1755 women were selected, but 3 women were subsequently disqualified leading to 585 women being assigned to the first and second cycles and 582 women being assigned to the final funding cycle.

[12] Figure 2.A.1, in Appendix 2.A, presents a timeline and sequence of activities for participants in the three funding cycles, including the timing of the surveys.

2.3.2 Checking the integrity of randomized design

We test the assumption that baseline characteristics are uncorrelated with treatment status by comparing the distribution of the baseline characteristics of participants. We make several comparisons that take into account the changing composition of the treatment and control groups as the program is progressively rolled-out. The results are presented in Table 2.2.

In panel A, we present summary statistics (mean and standard deviations) of variables that may be impacted by the program (expenditure, income, savings, asset ownership) or that may mediate its impact (household size, previous business experience, education). The baseline characteristics of the participants (and their households) are similar to those of other ultra-poor households in other regions of northern Kenya, which suggests that the findings of this study may be generalizable to ultra-poor women across northern Kenya (Merttens et al., 2013). Average monthly expenditure per capita is approximately PPP USD 33.96, which is well below the national poverty line. Approximately 70% of this expenditure is on food. Households are relatively large and have approximately 3.8 children on average, with less than 50% of children enrolled in school. Many households are food insecure, with children going to bed hungry at least 2 times a month. Households also own very little livestock: less than one Tropical Livestock Unit (TLU) per capita, well below the self-sufficiency threshold for mobile pastoralists in East African ASALs (McPeak & Barrett, 2001). [13] However, more than half of the participants report having some form of business experience, typically petty trade or the selling of livestock and livestock products.

In panel B, we present the t-tests of the null hypothesis of equality of means at baseline. These results indicate that randomization was successful in creating groups of individuals that are observationally identical, and in only one case can we reject the null hypothesis at the conventional 5% level. This conclusion is reinforced by the results of a F-test of the joint effect of these variables on treatment status, reported in panel C.

2.3.3 Spillover effects and program anticipation

Given the geographical proximity of individuals in the treatment and control groups, it is possible that control households use and benefit from the products and services offered by the businesses established by the treated households. We investigate three possible pathways for such influence: lower prices to consumers

[13] Tropical Livestock Unit (TLU) is a standardized unit, designed to measure the size of a mixed livestock herd: 1 TLU is equivalent to 1 head of cattle, 0.7 camels, 10 sheep/goats, or 2 donkeys.

TABLE 2.2: Summary statistics and balance checks for the treatment and control groups

Variable:	Monthly income per capita	Monthly expenditure per capita	Monthly food expenditure per capita	Monthly non-food expenditure per capita	Total savings per capita	TLU per capita	Durable asset index	Meals per day	Nights that child has gone to bed hungry in past week	Proportion of children in school	Household Size	# children	Married	Years of education	Business experience	Benefitting from HSNP	Participating in CARE VSLA
Panel A: Means and standard errors of variables at baseline.																	
$\bar{X}_A$	21.770	34.562	24.182	10.380	3.772	0.683	-0.234	1.941	0.549	0.435	5.778	3.875	0.800	0.328	0.576	0.106	0.089
(standard error)	(0.925)	(1.516)	(1.188)	(0.747)	(0.344)	(0.030)	(0.169)	(0.016)	(0.027)	(0.012)	(0.079)	(0.071)	(0.017)	(0.060)	(0.020)	(0.013)	(0.012)
Observations	585	585	585	585	585	585	585	581	578	583	585	585	585	585	585	585	585
$\bar{X}_B$	22.319	34.480	23.862	10.617	3.920	0.640	0.113	1.950	0.576	0.442	5.692	3.737	0.831	0.470	0.562	0.103	0.106
(standard error)	(0.933)	(1.402)	(1.075)	(0.770)	(0.328)	(0.037)	(0.189)	(0.016)	(0.029)	(0.012)	(0.075)	(0.070)	(0.016)	(0.072)	(0.021)	(0.013)	(0.013)
Observations	585	585	585	585	585	585	585	585	579	579	585	585	585	585	585	585	585
$\bar{X}_C$	22.449	32.825	22.494	10.331	5.123	0.684	0.124	1.933	0.576	0.412	5.596	3.711	0.773	0.414	0.538	0.113	0.108
(standard error)	(0.995)	(1.215)	(0.874)	(0.648)	(0.598)	(0.034)	(0.179)	(0.014)	(0.029)	(0.011)	(0.077)	(0.070)	(0.017)	(0.070)	(0.021)	(0.013)	(0.013)
Observations	582	582	582	582	582	582	582	580	572	579	582	582	582	582	582	582	582
Panel B: t test comparison of means of baseline characteristics.																	
$H_0 : \bar{X}_A = \bar{X}_{B+C}$ (p-values)	0.593	0.610	0.466	0.916	0.123	0.540	0.099*	0.994	0.430	0.588	0.163	0.083*	0.919	0.145	0.302	0.899	0.220
$H_0 : \bar{X}_B = \bar{X}_C$ (p-values)	0.927	0.373	0.323	0.776	0.078*	0.379	0.967	0.426	0.991	0.075*	0.373	0.798	0.014**	0.575	0.411	0.551	0.901
$H_0 : \bar{X}_A = \bar{X}_C$ (p-values)	0.617	0.372	0.253	0.961	0.051*	0.307	0.146	0.711	0.496	0.171	0.100	0.102	0.264	0.351	0.192	0.685	0.268

Panel C: F-test from regression of treatment on variables above.[a]

Treatment group	Control group	F-Stat	p-value
A	B and C	0.76	0.723
B	C	1.18	0.283
A	C	1.15	0.308

Note: All monetary values are reported in 2014 USD, PPP terms. [a]Monthly food and non-food expenditure per capita are excluded from this regression.*, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively.

due to higher competition from new businesses; lower profits of non-REAP businesses due to increased competition from new businesses; and, easier access to loans, given higher savings.

Given that more than 95% of the businesses that are established by the treated individuals are in petty trade (primarily of food items), the main impact of increased competition among businesses, may be a consequent reduction in market prices. Although this reduction is not expected to be substantial given the large number of pre-existing businesses in each location, we are able to control for this general equilibrium effect through the inclusion of the number of pre-existing businesses as a control variable when estimating the effect of the program. [14]

A different path through which businesses started by REAP participants may affect the welfare of non-participant households is through a reduction in income from petty trade. We test for this possibility by examining the income from petty trade earned by participants. In Table 2.3 we report the average income from non-REAP petty trade for participants in groups *A*, *B* and *C* at baseline and endline. We find that income from petty trade decreases among those still waiting to join the program, i.e. group *C*, by approximately 8% but this decrease is not statistically significant. [15]

TABLE 2.3: Income from non-REAP petty trade at baseline and endline

	Overall	Group *A*	Group *B*	Group *C*
$\bar{X}_{baseline}$	0.767	0.443	0.887	0.957
(standard error)	(0.122)	(0.094)	(0.153)	(0.311)
Observations	1651	534	556	561
$\bar{X}_{endline}$	1.257	1.355	1.542	0.880
(standard error)	(0.182)	(0.367)	(0.382)	(0.150)
Observations	1651	534	556	561
$H_0 : \bar{X}_{baseline} = \bar{X}_{endline}$ (p-values)	0.026**	0.016**	0.112	0.824

*, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively.

Another potential source of spillover effects might be easier access to loans. Although only REAP participants can actively participate in all saving groups' activities, loans can be (and typically, are) extended to other members of the community, so that they can deal with shocks and emergencies (usually, health, or school and food related expenditures). We capture information on borrowing

[14] Overall, there were 1932 businesses before the program. The program funded 195 businesses (approximately 10% of the pre-existing number) in each funding cycle. See Table 2.B.1, in Appendix 2.B, for further details.

[15] Participants in groups *A* and *B*, however, experience large (and in the case of group *A*, statistically significant) gains in income from their non-REAP petty trade businesses, suggesting that they are able to transfer the training, experience and income gained from participating in REAP to their own individually-owned businesses.

from REAP SGs for all women, and therefore can control for this effect when estimating the impact of the program.

Finally, bias could arise from participants changing their behavior in anticipation of receiving the program. [16] It is unclear in which direction such an anticipation effect may bias our results. For example, if participants delay investments in anticipation of receiving the grants, then estimated treatment effects would be upward biased. On the other hand, treatment effects may be downward biased if participants awaiting the grant are more willing to invest given the certainty of receiving the grant, which may act as a form of insurance (Bianchi & Bobba, 2013).

The design of the study does not allow for a straightforward way to test for anticipation effects. However, if anticipation results in either behavior then we would expect to find differences between individuals that enroll in the program in the second and third funding cycles, given that one group would anticipate receiving funding six months sooner than the other. If this intuition is correct then these differences would be ascertainable during the midline survey (when group B would immediately receive the first grant while group C would still be six months away from participating in the program). [17] We check for differences in monthly income per capita, monthly expenditure per capita, monthly consumption per capita, savings per capita, TLU per capita, durable asset index, and the nights that a child has gone to bed hungry in the last week, our outcome variables, and find no statistically significant differences between groups *B* and *C*, as shown in Table 2.4. [18]

We also collected information on income earned from other businesses (besides REAP businesses) in all rounds of data collection, which allows us to examine if anticipation of the program led to investment in a business that did not exist at baseline. We find no evidence of statistically significant differences in income from own business between groups *B* and *C* at midline (not reported). This is not surprising given that we also find no statistically significant differences in how these two groups of participants allocate their time at midline (see panel A of Table 2.E.1 in Appendix 2.E) or in the proportion of women that have ever taken a loan. [19]

[16] Given that we are dealing with the ultra-poor it is difficult to conjecture how behavior would change in anticipation of this program. Individuals might try to observe other businesses and how they operate or business groups might meet to discuss what will happen when they are enrolled in the program, but both capital access and human capital constraints are likely to prevent them taking any action that would affect measured outcomes.

[17] The implicit assumption here is that although all individuals in group B and group C learn when they will join the program at the same time, they only change their behaviour in anticipation of the program closer to their enrolment date.

[18] These outcome variables are defined in Appendix 2.C and Appendix 2.D.

[19] Approximately 24.2% (24.3%) of group *B* (*C*) participants have ever taken a loan from banks, MFIs, moneylenders, savings and self-help groups, or family. More than 36% of participants in group *A* have accessed loans and this is statistically significantly higher compared to groups *B* and *C*.

TABLE 2.4: Testing for anticipation effects

Variable:	Monthly income per capita	Monthly expenditure per capita	Monthly consumption per capita	Total savings per capita	TLU per capita	Durable asset index	Nights that child has gone to bed hungry in past week
Panel A: Means and standard errors of outcome variables for participants in groups B and C.							
$\bar{X}_B$	26.263	49.906	47.418	3.683	1.031	2.014	0.463
(standard error)	(2.198)	(1.949)	(1.463)	(0.408)	(0.052)	(0.265)	(0.058)
Observations	566	566	566	566	566	566	546
$\bar{X}_C$	23.437	51.703	48.685	3.178	1.119	2.086	0.565
(standard error)	(1.354)	(2.191)	(1.546)	(0.544)	(0.057)	(0.272)	(0.044)
Observations	567	567	567	567	567	567	545
Panel B: t test comparison of means of outcome variables for participants in groups B and C.							
$H_0 : \bar{X}_B = \bar{X}_C$ (p-values)							
	0.274	0.540	0.552	0.458	0.254	0.848	0.163

Panel C: F-test from regression of treatment on variables above.

F-Stat	p-value
0.98	0.447

Note: All monetary values are reported in 2014 USD, PPP terms.

More than 90% of loans taken by women in groups *B* and *C* are used to purchase food, with less than 2% of the loans used for investment in a business or livestock while the remainder are mainly used to pay for medical emergencies and school fees. The limited use of loans for investment in businesses can be attributed to the limited access to capital in this region: Osterloh and Barrett (2007), for example, show that the average size of loans available in similar locations in northern Kenya are often not sufficient to cover the cost of transport to sites where provisions can be purchased.

Although these investigations are not sufficient to definitely disprove the possibility of anticipation effects, together they point to their limited importance, if any.

2.4 Main results

The random assignment of treatment status allows us to obtain unbiased estimates of the impact of REAP, and its variance (that takes into account stratification) by estimating the following regression for each outcome of interest:

$$Y_i(t) = \theta + \beta T_{ij}(t) + \delta Y_i(0) + \tau M_i + \varphi X_i(t) + \epsilon_i \qquad t, j = \{1, 2\} \tag{2.1}$$

where $Y_i(t)$ is the outcome of interest for household i, at time t (=1 if midline, and =2 if endline), $Y_i(0)$ is the baseline value of the outcome variable for household i, M_i is a set of sub-location dummy variables, and $X_i(t)$ is a matrix of control variables (including a dummy variable to indicate if an individual has ever borrowed from a REAP SG, the number of REAP businesses in an individual's sub-location and the number of non-REAP businesses in an individual's location). [20] Finally $T_{i\bullet}$ is treatment status of individual i.

Given the structure of the program, we can consider two sets of interventions: business training, a cash grant of USD 100, and mentoring, which are introduced first, and that we label as $(T_{\bullet 1})$ and are followed by savings training, an additional cash grant of USD 50 and continued mentoring, that we label as $(T_{\bullet 2})$. Simplifying notation, by dropping the i-th individual subscript, it is clear from the description of the program (and from Figure 2.2) that we can observe T_1 at both midline and endline ($T_1(1)$ and $T_1(2)$), and the joint effect of the two sets of interventions at the endline $(T_1(1) + T_2(2))$.

To estimate the impact of T_1 at $t = 1$ we use the data collected during the midline survey to compare group *A* to a combined control group formed by those benefiting from the program in the second and third cycles (i.e. groups *B* and

[20] Stratification took place at the sub-location level (77 sub-locations).

C). We refer to this impact as $\beta(T_1(1))$. We can similarly estimate the impact of T_1 on group B at $t = 2$ by using the endline data to compare group B to control group *C*. We refer to this impact as $\beta(T_1(2))$. We can then use these two estimates of impacts to test the hypothesis that the impact of T_1 is constant throughout the period:

$$H_0 : \beta(T_1(1)) = \beta(T_1(2)) \tag{2.2}$$

Failure to reject (2.2) would suggest that the impact of this subset of interventions is stable, providing further support to our assumption that there were no adverse effects from late entry into treatment (due, for example, to increased market competition).

It is important to notice that failure to reject (2.2) is not enough to plausibly identify the impact of T_2 in isolation given that, at the end of $t = 1$, beneficiaries of T_1 will potentially be different from the same individuals at $t = 0$ both in ways that are easy to control (asset ownership, for example) and in ways that are not easy to observe (experience in managing a business as part of a group, for example). Hence, without further assumptions regarding how such variables influence the outcomes we analyze, we can only identify the effect of T_2 conditional on previously benefiting from T_1. To do that, we use the endline data to estimate the combined impact of T_1 and T_2 at $t = 2$, $[\beta(T_1(1) + T_2(2))]$, by comparing group A with control group *C*.

2.4.1 The six month impact of REAP

Table 2.5, panel A, provides the estimates of the impact of T_1 in both periods. Asterisks denote statistical significance based on the unadjusted p-values but we also adjust p-values (reported in brackets) to account for multiplicity. Because we estimate the impacts of REAP on several outcomes, some outcomes may display significance even if no effect exists since we have increased the probability of type 1 errors by testing multiple simultaneous hypotheses at set p-values. [21] Several methods exist to adjust p-values for multiple-inference and in this study we implement the step-up method to control for the false discovery rate (FDR) as proposed by Benjamini and Hochberg (1995). Using the procedure outlined by Anderson (2008) we are able to obtain adjusted p-values or q-values, which should be interpreted as the smallest significance level at which the null hypothesis is rejected.

[21] By performing seven independent tests, the probability of a type 1 error is no longer 0.05 but instead 0.302.

TABLE 2.5: The impacts of REAP on household outcomes

Outcome:	Monthly income per capita		Monthly expenditure per capita		Monthly consumption per capita		Total savings per capita		TLU per capita		Durable asset index		Nights that child has gone to bed hungry	
	$t=1$	$t=2$	$t=1$	$t=2$	$t=1$	$t=2$	$t=1$	$t=2$	$t=1$	$t=2$	$t=1$	$t=2$	$t=1$	$t=2$
Panel A: Estimates of the impact of $T_1(t)$														
$\hat{\beta}(T_1(t))$	11.276***	8.238***	5.079	0.819	2.802	-0.822	1.095	1.666**	-0.016	0.205**	0.379	0.743*	-0.103*	-0.184**
	(2.822)	(2.790)	(3.417)	(4.015)	(2.533)	(2.089)	(0.573)	(0.739)	(0.060)	(0.100)	(0.333)	(0.399)	(0.059)	(0.080)
	[0.001]	[0.024]	[0.241]	[0.839]	[0.314]	[0.810]	[0.194]	[0.058]	[0.795]	[0.074]	[0.314]	[0.089]	[0.194]	[0.058]
Observations	1682	1117	1682	1117	1682	1117	1682	1117	1682	1117	1682	1117	1597	1089
R-squared	0.107	0.110	0.280	0.102	0.222	0.284	0.116	0.105	0.331	0.224	0.318	0.408	0.260	0.089
Control group mean	24.849	25.232	50.805	57.394	48.052	46.245	3.430	4.440	1.075	1.303	2.050	2.843	0.514	0.789
Panel B: Estimates of the impact of $T_1(1)+T_2(2)$.														
$\hat{\beta}(T_1(1)+T_2(2))$		8.589***		-3.509		-2.542		5.832***		0.163		0.814**		-0.170**
		(2.232)		(3.453)		(2.582)		(0.789)		(0.107)		(0.368)		(0.084)
		[0.001]		[0.326]		[0.326]		[0.001]		[0.178]		[0.049]		[0.049]
Observations		1095		1095		1095		1095		1095		1095		1068
R-squared		0.127		0.126		0.252		0.148		0.272		0.424		0.103
Control group mean		25.232		57.394		46.245		4.440		1.303		2.843		0.789

Note: Regressions include controls for loans taken from REAP savings groups, number of REAP businesses in a manyatta, non-REAP businesses in a location, sub-location fixed effects and baseline levels of the outcome variable (with the exception of monthly consumption for which baseline levels are not available). Robust standard errors, clustered at the business group level, are shown in parentheses, while q-values, using the Benjamini-Hochberg step-up method, are shown in brackets. All monetary values are reported in 2014 USD, PPP terms. *, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively (based on unadjusted p-values).

After accounting for the possibility of simultaneous inference (by adjusting p-values), and searching for consistent impacts across all periods, we can only conclude that, after six months of benefiting from REAP, beneficiaries have higher income per capita. These changes are economically significant in both periods, and they represent an improvement of 45.4% over the control group mean (or 0.260 SDs) at $t = 1$ and 32.6% over the control group mean (or 0.236 SDs) at $t = 2$.

However, and somewhat surprisingly, these changes do not seem to translate into changes in monthly expenditure per capita which, although positive, are much less precisely estimated. This is especially true during $t = 2$, when we can reject the equality between increases in income and expenditure (p-value=0.048). [22]

One explanation for this discrepancy is that additional income is being allocated to asset accumulation rather than consumption. Our data offers some support to this explanation, in particular for $t = 2$, during which we observe a negligible (and statistically insignificant) decrease in consumption, an increase in savings and assets (both livestock and other assets) and a reduction in the number of nights a child has gone to bed hungry. Despite this apparent difference in the impact of T_1 between periods, with the effects being generally more positive in the second period, we can never reject the null hypothesis of equality of impact across periods (equation 2.2). [23]

Limiting our discussion to the changes identified in $t = 2$, we can conclude that, as with income per capita, changes in wealth (savings and assets) are economically important: per capita savings are 37.5% higher among compared beneficiaries (or 0.220 SDs), while durable asset ownership is higher by 26.1% (or 0.111 SDs). Finally, livestock ownership is also significantly higher in the second period (at the 10% level) with participants in the treatment group owning 15.7% (or 0.128 SDs) more livestock per capita compared to the control group. We discuss the possible reasons for the differences across periods after the analysis of the one year impact of the program, to which we now turn.

2.4.2 The one year impact of REAP

Table 2.5, panel B provides estimates of the combined impact of T_1 and T_2 (i.e. $\hat{\beta}(T_1(1) + T_2(2))$, after one year of participation in REAP. These estimates are in line with the ones presented in panel A, (i.e. the impact of T_1), with treated participants reporting significantly higher income per capita, savings per capita, and asset ownership. After one year of participation in REAP, income per capita is 34.0% (0.246 SDs) higher compared to the control group mean and savings per

[22] However, we cannot reject this equality during $t = 1$ at the usual levels of significance (p-value=0.119).

[23] Depending on outcome, the q-values are between 0.387 and 0.537. Specific results are available from the authors on request.

capita is 131.4 % (0.769 SDs) higher compared to the control group mean, with both increases statistically significant at the 5% level of significance.

As before, we find that the increase in household income does not translate to an increase in expenditure or consumption, which in fact decrease by 6.1% (0.061 SDs) and 5.5% (0.074 SDs) respectively, although these decreases are not statistically significant. [24] We find a similar impact on livestock and durable asset ownership at one year compared to six months, with both outcomes increasing as a result of REAP. The impact of REAP on the durable asset index represents a 28.6% (0.122 SDs) increase over the control group mean, and the impact on livestock represents a 12.5% (0.102 SDs) increase over the control group mean. However, only the increase in the durable asset index is statistically significant (at the 10% level). The estimates in Table 2.5 also reveal that participation in REAP results in a decrease in the instances in which a child is reported as going to bed hungry in the past week, a decrease that is statistically significant at the 10% level and represents a 21.5% (0.141 SDs) decrease compared to the control group mean.

Since T_2 is never implemented in isolation, we can only estimate its impact conditional on the implementation of T_1. As argued above, treated individuals may have changed in ways that are different to control individuals (experience in managing a business as part of a group, for example), making the impact of the second set of interventions unidentifiable without further assumptions.

We find that T_2 has a positive and statistically significant impact on savings per capita, with participants saving 106.7% more compared to the control group mean (Table 2.6). This impact is expected since one of the interventions in T_2 provides training on savings and helps participants to establish savings groups. We do not find any significant impacts on other outcomes of interest after adjusting for FDR.

2.4.3 Discussion

Income. The Rural Entrepreneur Access Project significantly increased the income earned by participants in the short-to-medium-run (i.e., 6 months and 1 year after participation in the program). The obvious mechanism through which the program may have led to this outcome is the formation of new micro-enterprises. One important question is whether such new enterprises crowd-out existing sources of income.

The results presented in Table 2.7 directly address this question by disaggregating income changes by source. The first conclusion is that the overall increase in income is being driven by changes in income from non-agricultural trade,

[24] Such decreases in expenditure and consumption are not unusual to observe in the short-to-medium run. For example, Keswell and Carter (2014) find that up to 12 months after benefiting from a transfer of land, households in South Africa experience a dip in consumption, which then rises to 150% of their pre-transfer level after three to four years

TABLE 2.6: The estimated impact of T_2 conditional on the participant receiving T_1

	$\hat{\beta}(T_1(1))$	$\hat{\beta}(T_1(1)+T_2(2))$	$\hat{\beta}(T_1(1)+T_2(2))$ $-\hat{\beta}(T_1(1))$	q-value for all 6 hypotheses
Monthly income per capita	11.276*** (2.822)	8.589*** (2.232)	-2.687 [0.398]	0.465
Monthly expenditure per capita	5.079 (3.417)	-3.509 (3.453)	-8.588** [0.048]	0.112
Monthly consumption per capita	2.802 (2.533)	-2.542 (2.582)	-5.345** [0.030]	0.104
Total savings per capita	1.095* (0.573)	5.832*** (0.789)	4.737*** [0.000]	0.001***
TLU per capita	-0.016 (0.060)	0.163 (0.107)	0.179* [0.094]	0.164
Durable asset index	0.379 (0.333)	0.814** (0.368)	0.435 [0.207]	0.290
Nights that child has gone to bed hungry in past week	-0.103* (0.059)	-0.170** (0.084)	-0.067 [0.490]	0.490

Note: In columns (1) and (2) robust standard errors reported in parentheses. In column (3) p-values from a Wald test of the null hypothesis $H_0 : \hat{\beta}(T_1(1) = \hat{\beta}(T_1(1) + T_2(2))$ are reported in squared brackets. In column (4) we estimate q-values based on the p-values reported in column (3).
*, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively (based on unadjusted p-values).

which includes income from the REAP microenterprise (recall that more than 95% of groups invest in petty trade businesses). The increase in income from non-agricultural trade is statistically significant at the 5% level of significance and this effect persists for up to one year after being enrolled in REAP. The second conclusion is that increased business activity does not crowd out other sources of income, suggesting that the program is bringing idle resources into productive activities. When we examine how participants allocate their time resources at $t = 2$ we find that those that have benefited from REAP are spending approximately 6% of their day on REAP related activities on average, and to achieve this increased activity they have decreased the average time spent on leisure and household activities, as well as other productive activities (see Table 2.E.2 in Appendix 2.E). [25]

It should be noted that the increase in income from non-agricultural trade is significantly lower in $t = 2$ for both treatment groups compared to $t = 1$. This result points to the importance of seasonality in the evaluation of this program, with the fluctuation in income from $t = 1$ to $t = 2$ likely due to seasonality in production in the region. This is supported by the fact that the total value of the business (i.e. the sum of cash on hand, business savings and credit outstanding, and business stock and assets) is significantly higher at $t = 2$ (for both sets of participants) compared to the business value at $t = 1$ (PPP USD 374.61 and PPP USD 451.55 for the six month and one year groups at $t = 2$, respectively, compared to PPP USD 305.50 for the six month group at $t = 1$), despite significantly lower incomes from non-agricultural trade at $t = 2$.

[25]The income generated from other productive activities has not declined in either treatment group at $t = 2$.

TABLE 2.7: The impact of REAP on various sources of income

Variable:	Monthly total income per capita	Monthly income from livestock per capita	Monthly income from other agriculture per capita	Monthly income from non-agri trade per capita	Monthly income from labor per capita	Monthly income from transfers per capita
Panel A: The impact at six months measured at $t = 1$						
$\hat{\beta}(T_1(1))$	11.276***	1.368	-0.013	9.863***	-0.085	0.005
	(2.822)	(2.462)	(0.069)	(0.976)	(0.301)	(0.223)
Observations	1682	1682	1682	1682	1682	1682
R-squared	0.107	0.091	0.123	0.288	0.162	0.108
Control group mean	24.849	19.817	0.117	3.031	1.205	0.678
Panel B: The impact at six months measured at $t = 2$						
$\hat{\beta}(T_1(2))$	8.238***	3.649	0.109	4.480***	0.192	-0.156
	(2.790)	(2.432)	(0.102)	(0.589)	(0.951)	(0.261)
Observations	1117	1117	1117	1117	1117	1117
R-squared	0.110	0.118	0.135	0.144	0.073	0.104
Control group mean	25.232	19.811	0.172	1.534	2.911	0.803
Panel C: The impact at one year measured at $t = 2$						
$\hat{\beta}(T_1(1) + T_2(2))$	8.589***	1.602	0.077	5.853***	0.987	-0.062
	(2.232)	(1.810)	(0.099)	(0.772)	(1.093)	(0.277)
Observations	1095	1095	1095	1095	1095	1095
R-squared	0.127	0.121	0.172	0.174	0.128	0.100
Control group mean	25.232	19.811	0.172	1.534	2.911	0.803

Note: Regressions include controls for loans taken from REAP savings groups, number of REAP businesses in a manyatta, non-REAP businesses in a location, sub-location fixed effects and baseline levels of the outcome variable. Robust standard errors, clustered at the business group level, are shown in parentheses. All monetary values are reported in 2014 USD, PPP terms.
*, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively (based on unadjusted p-values).

Expenditure and consumption. We do not find any significant impacts of REAP on expenditure or consumption but in Table 2.6 we show that the effect on these outcomes are lower after one year of participation compared to after six months. Recall that one of the roles of mentors is to promote practices that would lead to successful businesses. Also recall that after six months of participation in REAP more than 95% of participants join savings groups where they are required to deposit savings on a monthly basis. These two factors are likely to result in the observed dip in consumption and expenditure after one year in REAP as participants may choose to divert additional income to savings and their businesses instead of additional consumption. [26]

Savings. As previously mentioned, after six months of participation in the program participants receive training on savings, including on the functioning of savings groups. After this training, more than 95% of participants join a SG, a decision that is both voluntary and individual (while at baseline only 10% were members of pre-existing SGs). It is therefore not surprising that after one year of participation in REAP, participants have saved more per capita.

What might be surprising is that we also find that before the training on savings, participants have also saved more per capita. This points to a shift in savings behavior that takes place even before the formal introduction of savings groups. If we look more closely at the savings mechanisms used by women (Table 2.8) we see that after six months REAP participants are saving more at home compared to the control group.

Livestock and other assets. Average livestock ownership among both the treatment and control groups has increased from baseline (0.669 TLU per capita) to midline (1.070 TLU per capita) to endline (1.405 TLU per capita), and, given the economic and social importance of livestock among participants, one would expect some of the increased income from entrepreneurial activities to be invested in the acquisition of livestock. We do find increased livestock ownership among REAP participants, which is in line with our expectations. By providing participants with an alternative source of income, REAP enables households to increase their herd size which is essential for pastoralist households to escape the poverty trap and to be able to recover from shocks that can push them back into poverty (Little et al., 2008), providing further evidence of how REAP can lead to sustained increases in well-being and graduate participants from ultra-poverty. Treated households also invest more in durable assets such as blankets, mosquito nets and latrines, which improve the living conditions of their households.

[26] See, for example, M. R. Carter, Tjernström, and Toledo (2016) who study a business development program in Nicaragua and similarly find no impact on consumption expenditures in the long run despite large increases in income. They attribute this finding on consumption to liquidity constrained households deciding to reinvest income increases rather than consume them.

TABLE 2.8: The impact of REAP on savings held using various mechanisms.

Variable:	Personal savings percapita	Non-BOMA savings group savings per capita	BOMA savings group savings per capita
Panel A: The impact at six months measured at $t = 1$			
$\hat{\beta}(T_1(1))$	1.038*	0.055	-
	(0.568)	(0.174)	-
Observations	1682	1682	-
R-squared	0.119	0.098	-
Control group mean	3.030	0.400	0
q-value	0.136	0.750	-
Panel B: The impact at six months measured at $t = 2$			
$\hat{\beta}(T_1(2))$	1.624***	-0.010	-
	(0.552)	(0.170)	-
Observations	1117	1117	-
R-squared	0.107	0.124	-
Control group mean	4.082	0.357	0
q-value	0.006***	0.952	-
Panel C: The impact at one year measured at $t = 2$			
$\hat{\beta}(T_1(1) + T_2(2))$	1.450***	0.246	4.141***
	(0.544)	(0.303)	(0.131)
Observations	1095	1095	1095
R-squared	0.087	0.063	0.605
Control group mean	4.082	0.357	0
q-value	0.012**	0.417	0.001***

Note: Regressions include controls for loans taken from REAP savings groups, number of REAP businesses in a many-atta, non-REAP businesses in a location, sub-location fixed effects and baseline levels of the outcome variable. Robust standard errors, clustered at the business group level, are shown in parentheses. All monetary values are reported in 2014 USD, PPP terms. Personal savings includes savings kept at home and savings kept at a formal financial institution including mobile service providers.

*, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively (based on unadjusted p-values).

Graduation from poverty. The main aim of this program is to graduate participants from poverty, which we equate with being above the Kenya rural poverty line as reported by the Kenya National Bureau of Statistics (2007). In Table 2.9 we provide estimates of the impact of REAP on the probability of being non-poor at six months and one year after the start of the program, when poverty lines are defined in terms of income or expenditure.

We find that beneficiaries are more likely to have incomes above the poverty line both after six months and one year of participation in REAP, and these effects are statistically significant at the 1% level. At $t = 1$ ($t = 2$) we find that T_1 increases the probability that beneficiaries are above the poverty line by 12.6% (6.6%), an effect that represents a 74.3% (39.6%) increase over the control group probability of being above the poverty line. The effects are similar at one year, with beneficiaries being 12.9% more likely to have incomes above the poverty line (a 77.0 % increase over the control group). When looking at the impact on the probability that a beneficiary has expenditure or consumption above the poverty line we find a slight increase in the treated group at $t = 1$ and a slight decrease at $t = 2$. However, none of these impacts are statistically significant at conventional levels, as

TABLE 2.9: The impact of REAP on the probability of living above the poverty line.

Variable:	Proportion of households with monthly income per adult equivalent above the poverty line		Proportion of households with monthly expenditure per adult equivalent above the poverty line		Proportion of households with monthly consumption per adult equivalent above the poverty line	
Panel A: Estimates of the six month impact at $t = 1$ and $t = 2$.						
	$t = 1$	$t = 2$	$t = 1$	$t = 2$	$t = 1$	$t = 2$
$\hat{\beta}(T_1(t))$	0.126***	0.066***	0.034	-0.003	0.009	-0.016
	(0.024)	(0.025)	(0.027)	(0.032)	(0.026)	(0.030)
Observations	1682	1117	1682	1117	1682	1117
R-squared	0.179	0.092	0.335	0.113	0.343	0.257
Control group mean	0.169	0.168	0.500	0.581	0.591	0.526
Panel B: $H_0 : \hat{\beta}(T_1(1)) = \hat{\beta}(T_1(2))$						
	0.082*		0.365		0.524	
Panel C: Estimates of the one year impact.						
		$t = 2$		$t = 2$		$t = 2$
$\hat{\beta}(T_1(1) + T_2(2))$		0.129***		-0.017		-0.049
		(0.027)		(0.032)		(0.032)
Observations		1095		1095		1095
R-squared		0.123		0.102		0.264
Control group mean		0.168		0.581		0.526

Note: Estimates from a linear probability model. Regressions include controls for loans taken from REAP savings groups, number of REAP businesses in a manyatta, non-REAP businesses in a location, sub-location fixed effects and baseline levels of the outcome variable (with the exception of monthly consumption for which baseline levels are not available). Robust standard errors, clustered at the business group level, are shown in parentheses.
The Kenya rural poverty line used is as defined by the Kenya National Bureau of Statistics (2007) which, after conversion, is estimated to be 77.069 USD per month and per adult equivalent in PPP 2014 terms.

*, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively (based on unadjusted p-values).

expected, given the earlier findings on expenditure and consumption.

Impact Heterogeneity. We next consider the evidence for differentiated impacts of REAP across the distribution of outcomes. In Table 2.10 we present quantile regression estimates at the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of outcomes, at six months (panels A and B) and one year (panel C). In Figures 2.3a and 2.3b we graph the quantile regression estimates for each of the 99 percentiles of the distribution of outcomes, again distinguishing for the duration of participation in the program (six months vs. one year) and the two periods of data collection. [27] Taken together, these results suggest several conclusions.

The first is that the effects on income are positive and statistically significant at each of the five quantiles reported in Table 2.10, and these effects are increasing with the quantile of the distribution. [28] This is true for both time periods and irrespective of the length of participation in the program. Hence, it seems possible to conclude that REAP was particularly effective, in terms of increases in income and in the short-to-medium-run, for those who were better-off (relatively speaking, as we are still talking of extremely poor populations): the effect of the program estimated at the 90th percentile is almost four times the effect at the 10th percentile. If the motivation of the poverty graduation approach is to include the ultra-poor, we can then conclude that this approach may take longer (or require modifications) for those who are at the bottom of the distribution.

The second is that we also observe more pronounced effects among individuals in the upper quantiles of the other outcome distributions. These patterns are clearly illustrated in Figure 2.3b where we see larger treatment effects for those in the upper quantiles of the savings, livestock and durable asset distributions, particularly when these effects are measured at $t = 2$.

The third is that the timing of measurement of the impact of the program ($t = 1$ vs. $t = 2$) seems to matter more in terms of shaping the effect of the program than the length of exposure to the program (six months vs. one year), which likely reflects the importance of seasonality in the context we study. The exception to this conclusion is, clearly, savings for which we find evidence suggesting that the lack of access to savings institutions (or lack of awareness about their functioning) may have prevented individuals from keeping liquid savings. When these constraints

[27] Quantile regressions were estimated with the user-written command -qreg2- which allows for standard errors that are robust to intra-cluster correlation (Parente & Santos Silva, 2016). We were unable to reliably estimate quantile regressions for the outcome "number of nights that a child has gone to bed hungry in the past week", as this variable does not have a well-behaved density. We were also unable to estimate quantile regressions on savings per capita at $t = 1$ for the following percentiles: 0.02, 0.03, 0.04, 0.07, 0.08, 0.09, 0.11, 0.12, 0.13, 0.16, 0.19.

[28] The one exception is the six month effect (at $t = 2$) for the 50th percentile, which is not statistically significant at conventional levels.

TABLE 2.10: The quantile treatment effects of REAP

Outcome	OLS Estimates	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
Panel A: Treatment effects at six months (at $t = 1$)						
Monthly income per capita	11.276***	2.446***	5.041***	6.753***	10.266***	9.292***
	(2.822)	(0.690)	(0.933)	(1.264)	(1.997)	(3.274)
Monthly expenditure per capita	5.079	1.357	0.882	1.195	3.222	3.307
	(3.417)	(1.130)	(1.093)	(1.586)	(2.638)	(4.605)
Monthly consumption per capita	2.802	-0.183	0.560	0.811	2.963	4.564
	(2.533)	(0.967)	(0.954)	(1.378)	(2.255)	(3.647)
Total savings per capita	1.095*	0.000	0.000	-0.000	0.265***	0.919***
	(0.573)	(0.000)	(0.000)	(0.000)	(0.074)	(0.353)
TLU per capita	-0.016	0.011	0.007	0.000	0.000	-0.001
	(0.060)	(0.026)	(0.030)	(0.037)	(0.056)	(0.076)
Durable asset index	0.379	-0.000	0.428	0.322	0.998**	0.622
	(0.333)	(0.317)	(0.294)	(0.334)	(0.440)	(0.498)
Panel B: Treatment effects at six months (at $t = 2$)						
Monthly income per capita	8.238***	1.931**	2.560***	2.393	6.610**	11.889***
	(2.790)	(0.795)	(0.852)	(1.605)	(2.966)	(4.331)
Monthly expenditure per capita	0.819	0.437	0.347	-0.259	-5.305	-8.576
	(4.015)	(1.732)	(1.825)	(2.377)	(4.339)	(8.033)
Monthly consumption per capita	-0.822	1.593	0.364	0.480	–0.095	-1.009
	(2.089)	(1.172)	(1.128)	(1.368)	(2.122)	(3.545)
Total savings per capita	1.666**	-0.000	-0.008	0.384	0.897	3.552***
	(0.739)	(0.216)	(0.151)	(0.378)	(0.577)	(1.018)
TLU per capita	0.205**	0.055	0.097***	0.102	0.195**	0.359**
	(0.100)	(0.036)	(0.037)	(0.065)	(0.093)	(0.151)
Durable asset index	0.743*	-0.167	0.036	0.650	0.944**	2.033***
	(0.399)	(0.389)	(0.371)	(0.494)	(0.418)	(0.746)
Panel C: Treatment effects at one year (at $t = 2$)						
Monthly income per capita	8.589***	2.611***	4.210***	6.080***	10.558***	13.985***
	(2.232)	(0.985)	(1.113)	(1.485)	(3.316)	(5.318)
Monthly expenditure per capita	-3.509	-0.715	-1.960	-1.988	-6.689*	-1.503
	(3.453)	(1.720)	(1.669)	(2.496)	(3.564)	(6.110)
Monthly consumption per capita	-2.542	0.118	-2.329*	-2.698*	-4.660**	-0.723
	(2.582)	(1.148)	(1.398)	(1.467)	(2.043)	(3.545)
Total savings per capita	5.832***	2.744***	3.566***	4.826***	6.636***	8.139***
	(0.789)	(0.349)	(0.294)	(0.473)	(0.820)	(1.134)
TLU per capita	0.163	0.081*	0.097**	0.106**	0.037	0.070
	(0.107)	(0.042)	(0.044)	(0.050)	(0.111)	(0.131)
Durable asset index	0.814**	-0.000	0.330	0.548	0.851	2.005***
	(0.368)	(0.390)	(0.327)	(0.371)	(0.644)	(0.549)

Note: Regressions include controls for loans taken from REAP savings groups, number of REAP businesses in a manyatta, non-REAP businesses in a location, sub-location fixed effects and baseline levels of the outcome variable (with the exception of monthly consumption for which baseline levels are not available). Robust standard errors, clustered at the business group level, are shown in parentheses. All monetary values are reported in 2014 USD, PPP terms.

*, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively.

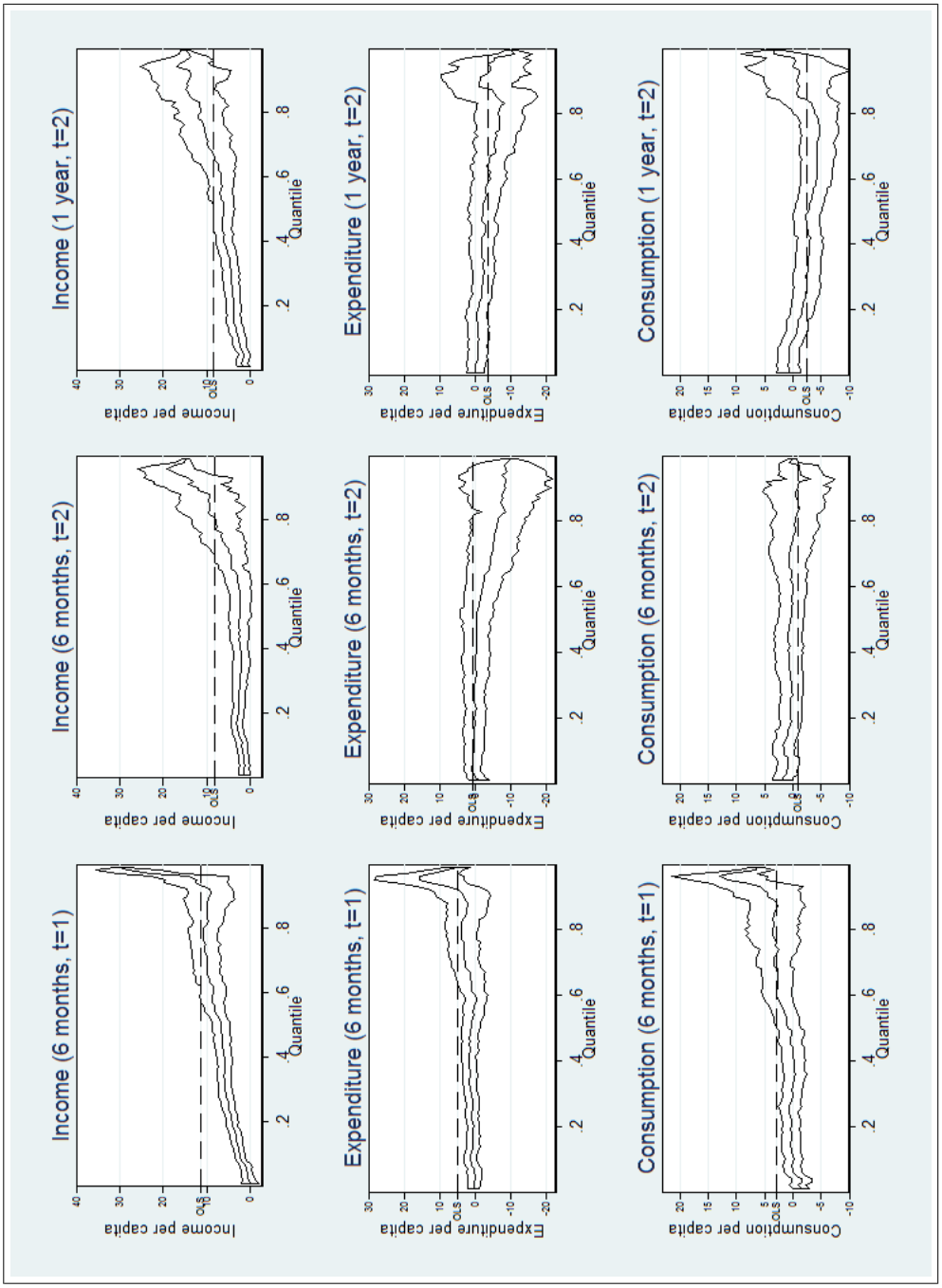


FIGURE 2.3A: The quantile treatment effects of REAP (income, expenditure and consumption).

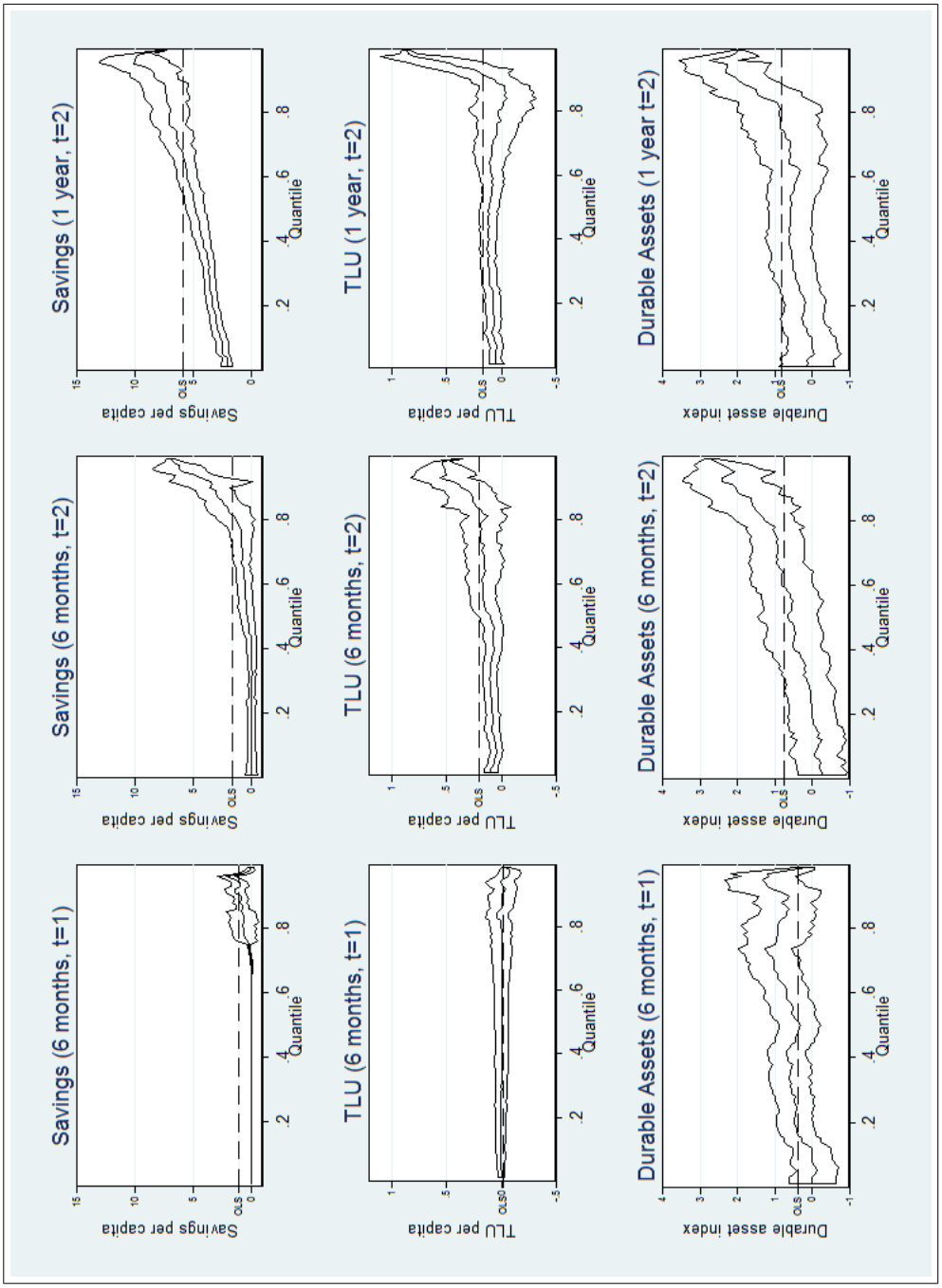


FIGURE 2.3B: The quantile treatment effects of REAP (savings, TLU and durable assets).

are removed (through the promotion of savings groups) we find significant treatment effects across the entire distribution and not just the upper quantiles. [29]

Finally, we would expect that those individuals with higher incomes (who gain most from REAP, in terms of income) would also be the ones who would show higher effects of participating in the program in terms of other variables such as savings or investment in livestock or other durables. The similarity in the patterns exhibited in Table 2.10 and Figures 2.3a and 2.3b could be thought to suggest some support to that expectation. To determine if this is true, we check whether individuals occupy similar quantile positions in the conditional distribution of income and of other outcome variables. In Table 2.11 we present the proportion of individuals who are in the 90th percentile of different combinations of outcome variables. It turns out that, for most pairs of outcome variables, less than 25% of individuals are in similar places in the distribution of outcomes. This result suggests that beneficiaries may employ different strategies, with some choosing to invest more in productive assets such as livestock, some opting for durable assets or liquid savings, and others choosing to consume. Such fundamental heterogeneity is reminiscent of the distinction between subsistence and transformative entrepreneurship (Schoar, 2009) but we leave a deeper analysis of these differences for future research.

Comparison of our findings to other studies. Finally, it seems also important to notice that our estimates of the impact of this program are of a similar order of magnitude to previous studies, namely Banerjee, Duflo, et al. (2015) and Bandiera et al. (2016). After one year, we find a 34% increase in income compared to the control group, similar to the increases in income that can be estimated from the results presented in Banerjee, Duflo, et al. (2015) and Bandiera et al. (2016). [30] The estimate of the impact of the program on savings (131.4% increase) is also similar to those estimated by Banerjee, Duflo, et al. (2015) who report a 155.5% increase after two years and 95.7% increase after three years. Our indicator of food security (number of nights that child has gone to bed hungry in the past week) is most similar to the variable "everyone in the household gets enough food everyday" reported on by Banerjee, Duflo, et al. (2015). They find that this variable improves by 10% (20%) after two years (three years) and we find a similar result, with our indicator improving by 21.5% after one year. Overall we find that REAP increases the probability of being above the poverty line by 12.9% which is similar to the 11% shift in women out of extreme poverty estimated by Bandiera et al. (2016). However, as with the other ultra-poor poverty graduation programs, our findings are more conservative compared to those of Blattman et al. (2016), which

[29] Note that before the introduction of savings groups, we only observe significant effects on savings in the upper quantiles (75th and 90th at $t = 1$, and 90th at $t = 2$) of the savings distribution.

[30] Banerjee, Duflo, et al. (2015) find an average increase of 25.7% (22.8%) across four sources of income after two years (three years), and Bandiera et al. (2016) find a 38% increase in income after four years.

TABLE 2.11: Overlap between individuals in the 90th percentile of the outcome distribution

Outcome	Monthly income per capita	Monthly expenditure per capita	Monthly consumption per capita	Total savings per capita	TLU per capita	Durable asset index
Panel A: Overlap at six months (at $t = 1$)						
Monthly income per capita	1	0.254	0.225	0.148	0.231	0.183
Monthly expenditure per capita		1	0.249	0.195	0.166	0.195
Monthly consumption per capita			1	0.254	0.225	0.154
Total savings per capita				1	0.112	0.225
TLU per capita					1	0.101
Durable asset index						1
Panel B: Overlap at six months (at $t = 2$)						
Monthly income per capita	1	0.286	0.107	0.169	0.214	0.250
Monthly expenditure per capita		1	0.295	0.143	0.134	0.205
Monthly consumption per capita			1	0.188	0.089	0.179
Total savings per capita				1	0.125	0.232
TLU per capita					1	0.134
Durable asset index						1
Panel C: Overlap at one year (at $t = 2$)						
Monthly income per capita	1	0.236	0.145	0.182	0.109	0.282
Monthly expenditure per capita		1	0.300	0.155	0.127	0.191
Monthly consumption per capita			1	0.291	0.136	0.173
Total savings per capita				1	0.109	0.236
TLU per capita					1	0.118
Durable asset index						1

Note: Figures represent the proportion of overlap between individuals in the 90th percentile of the two corresponding outcome distributions.

may be explained both by the post-war setting that they study and the possibility of anticipation effects, that cannot be ruled out in their study.

Turning to the cost-benefit analysis of this program, it is estimated that the cost for one additional woman to be enrolled in REAP in 2015 for two years was approximately USD 300 or PPP USD 713.91 at 2014 prices. This figure is well below the direct costs of the six programs evaluated by Banerjee, Duflo, et al. (2015) as well as the program evaluated by Blattman et al. (2016). Assuming that this was the cost to implement the program in 2013, and ignoring discounting and inflation, the gains in income (which we estimate to be the average of the one year and six month impacts) would have to persist for one additional month for the gains in income to cover the cost of the program.

2.5 Conclusion

In this paper we study a multifaceted approach to poverty alleviation that is being increasingly recognized for its ability to set ultra-poor households on a sustainable pathway out of extreme poverty (Bandiera et al., 2016; Banerjee, Duflo, et al., 2015).

We show that a variation of the BRAC approach, the Rural Entrepreneur Access Project (REAP), that provides disadvantaged women with capital, skills, and ongoing mentorship in business and savings, but that excludes consumption support, replaces asset transfers with cash transfers, and targets groups instead of individuals, has enabled beneficiaries to run microenterprises that led to improved household incomes. These short-to-medium-run impacts are economically significant and allow women to meet current household needs (through increased incomes) and plan for future shocks (through the accumulation of liquid savings and assets). The pathway of change is quite clear, with a tightly-estimated shift of time use from leisure and household activity into non-farm enterprise activity, with 95% of enterprises involved in petty trade of consumer goods.

The estimates of the impact of this program are, largely, in line with other evaluations of similar programs (Bandiera et al., 2016; Banerjee, Duflo, et al., 2015), and with a similar intervention examined by Blattman et al. (2016), with a relatively similar evaluation horizon. And, although the existing data do not allow us to examine the sustainability of the impacts once participants stop receiving support, the similarity in results between our analysis and prior ultra-poor trials raises a plausible prospect that these impacts should be stable over time. [31] A simple cost-benefit analysis shows that if this were true, the program would cover costs within a reasonable time horizon.

We are also able to demonstrate the potential for this approach to be applied in a different, arguably more extreme context to those already studied. The Rural Entrepreneur Access Project was implemented in some difficult to work in locations, with very low average population densities, highly variable weather conditions, low infrastructure, and limited access to markets. Yet, women were able to make use of the capital and skills delivered through REAP to establish and run successful enterprises. This consistency of results provides important support for the robustness of the poverty graduation approach and further corroborates the external validity of the other studies, while suggesting further opportunities for experimentation in the design and implementation of such programs.

[31] Banerjee, Duflo, et al. (2015) examine two year and three year impacts and find no evidence of mean reversion of the impacts. Bandiera et al. (2016) look at two year and four year impacts and find more pronounced effects across many outcomes after four years compared to after two years.

Appendix

2.A Timeline of activities

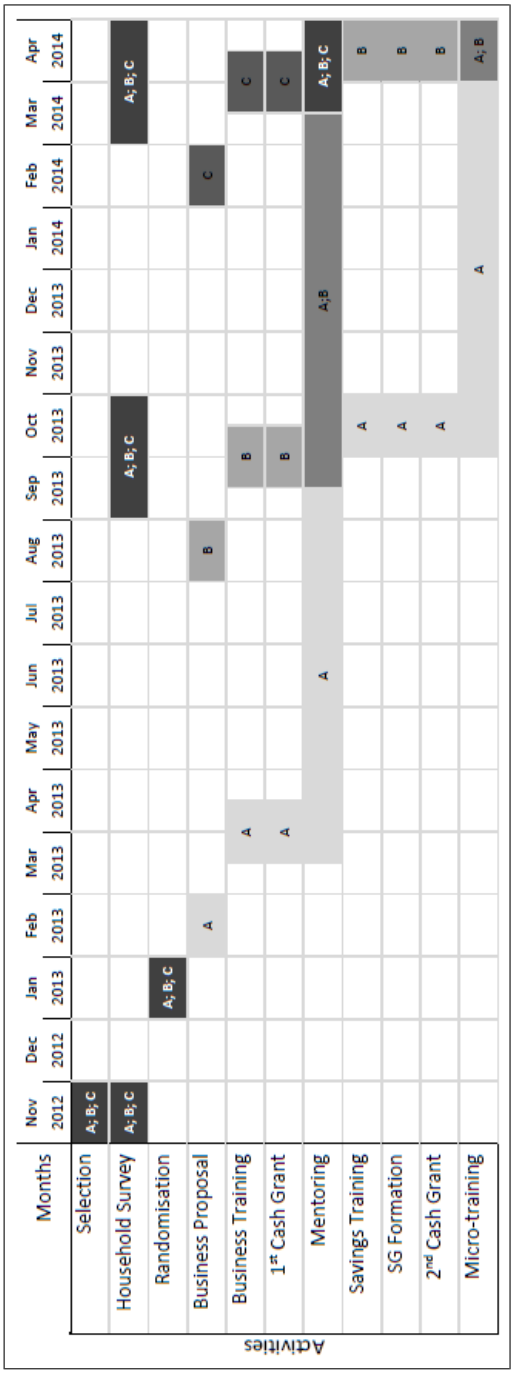


FIGURE 2.A.1: Timeline of data collection and program activities, by assignment to funding cycle.

Note: We label the beneficiaries who entered the program in the first cycle as group A, and subsequent groups as B and C.

2.B Pre-existing businesses

TABLE 2.B.1: Population and number of businesses by location.

Location	Population[a]	Pre-existent businesses	Businesses formed between March 2013 and April 2014[b]
1	13012	241	60
2	8357	159	30
3	7000	146	30
4	7800	227	30
5	4078	99	30
6	3300	70	30
7	10238	167	60
8	8935	131	60
9	4226	87	30
10	11220	289	60
11	3076	27	30
12	4065	56	30
13	8030	89	60
14	11223	144	44

Note: Information on the number of businesses was provided by the BOMA Project.

[a]Population numbers are based on the 2009 Kenya Census.

[b]One-third of new REAP businesses were formed in each round of funding with the exception of location 14 where 14 businesses were formed in March/April 2014 and 15 businesses in each of the two previous rounds.

2.C Outcome variables

TABLE 2.C.1: Description of outcome variables included in analysis.

Variable	Description
Monthly income per capita	Income from 1) REAP business, 2) non-REAP business, 3) livestock and livestock products, 4) firewood and charcoal, 5) water, 6) crops, 7) salaried and casual labour, 8) pension, transfers and remittances.
Monthly expenditure per capita	Expenditure on 1) food, 2) clothes, 3) school, 4) health, 5) household items, 6) household repairs, 7) livestock, 8) travel, 9) cosmetics, 10) beads, 11) ceremonies.
Monthly consumption per capita	Monetary value of consumption of food and fuel.
Total savings per capita	Savings held 1) at home, 2) with formal financial institutions and mobile service providers, 3) with non-BOMA savings groups, 4) with BOMA savings groups.
TLU per capita	1 TLU is equivalent to 1 head of cattle, 0.7 camels, 10 sheep/goats, or 2 donkeys.
Durable asset index	As defined in Appendix 2.D
Nights that child has gone to bed hungry in past week	The number of nights in the past week that any child in the household is reported as going to bed hungry.

2.D Durable asset index

Filmer and Pritchett (2001) were among the first to suggest the use of principle component analysis (PCA) to aggregate several asset ownership variables into a single dimension. Principle component analysis was seen as a more methodologically sound way of assigning weights to the variables that comprise an index compared to other methods, such as simple summation or the use of asset values. However, the use of PCA for this purpose has come under criticism since one of the assumptions underlying PCA is that variables are continuous and normally distributed which is violated when discrete variables are included in the analysis (Howe, Hargreaves, & Huttly, 2008). Multiple correspondence analysis (MCA) has been suggested as an alternative approach that is analogous to PCA but is better suited for use with discrete data (Booysen, van der Berg, Burger, von Maltitz, & du Rand, 2008).

We make use of the approach suggested by Booysen et al. (2008) to create an asset index including information on the ownership of 11 durable assets that were determined in all survey rounds. The assets include: 1) blanket, 2) flask, 3) kitchen,

4) lamp, 5) latrine, 6) mattress, 7) mobile phone, 8) mosquito net, 9) nylon sheet, 10) slasher, and, 11) spade. Using the –*mca*– command in Stata 13 we find that the first dimension accounts for 47% of the inertia.[32] We use the coordinates reported for the first dimension to generate weights for every asset included in the index. These weights are reported in Table 2.D.1.

TABLE 2.D.1: Variables and MCA weights used in asset index.

Asset	Category	Weight
Owns a blanket	No	-0.543
	Yes	0.296
Owns a flask	No	-0.883
	Yes	1.395
Owns a kitchen	No	-0.372
	Yes	2.905
Owns a lamp	No	-0.483
	Yes	3.994
Owns a latrine	No	-0.165
	Yes	5.612
Owns a mattress	No	-0.534
	Yes	4.662
Owns a mobile phone	No	-0.354
	Yes	4.343
Owns a mosquito net	No	-1.234
	Yes	0.600
Owns a nylon sheet	No	-0.452
	Yes	0.175
Owns a slasher	No	-0.336
	Yes	0.032
Owns a spade	No	-0.322
	Yes	2.540

[32] Inertia is Pearson's chi-squared divided by sample size and is analogous to variance reported on in PCA.

2.E Use of time

TABLE 2.E.1: Use of time at midline.

Variable:	Proportion of last day in leisure	Proportion of last day in household activities	Proportion of last day in REAP business activities	Proportion of last day in non-REAP business activities	Proportion of last day in other productive activities	Proportion of last day in social activities	Proportion of last day in other activities
Panel A: Means and standard errors of outcome variables for participants in groups A, B and C.							
$\bar{X}_A$	0.449	0.348	0.075	0.003	0.114	0.009	0.003
(standard error)	(0.003)	(0.001)	(0.004)	(0.001)	(0.004)	(0.001)	(0.001)
Observations	538	538	538	538	538	538	538
$\bar{X}_B$	0.482	0.382	0	0.011	0.113	0.008	0.004
(standard error)	(0.004)	(0.004)	(0)	(0.002)	(0.004)	(0.002)	(0.001)
Observations	525	525	525	525	525	525	525
$\bar{X}_C$	0.480	0.384	0	0.009	0.117	0.007	0.003
(standard error)	(0.004)	(0.004)	(0)	(0.001)	(0.004)	(0.001)	(0.001)
Observations	535	535	535	535	535	535	535
Panel B: t test comparison of means of outcome variables for participants in groups B and C.							
$H_0 : \bar{X}_B = \bar{X}_C$ (p-values)	0.786	0.779	.	0.251	0.543	0.802	0.553
Panel C: t test comparison of means of outcome variables for participants in groups A and C.							
$H_0 : \bar{X}_A = \bar{X}_C$ (p-values)	0.000***	0.000***	0.000***	0.000***	0.598	0.224	0.749
Panel D: t test comparison of means of outcome variables for participants in groups A and B.							
$H_0 : \bar{X}_A = \bar{X}_B$ (p-values)	0.000***	0.000***	0.000***	0.000***	0.928	0.367	0.790

Note: Eighty-four use of time observations are missing for midline survey respondents. *, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively.

TABLE 2.E.2: Use of time at endline.

Variable:	Proportion of last day in leisure	Proportion of last day in household activities	Proportion of last day in REAP business activities	Proportion of last day in non-REAP business activities	Proportion of last day in other productive activities	Proportion of last day in social activities	Proportion of last day in other activities
Panel A: Means and standard errors of outcome variables for participants in groups A, B and C.							
$\bar{X}_A$	0.416	0.374	0.058	0.004	0.123	0.002	0.023
(standard error)	(0.003)	(0.004)	(0.003)	(0.001)	(0.004)	(0.001)	(0.003)
Observations	533	533	533	533	533	533	533
$\bar{X}_B$	0.415	0.381	0.056	0.003	0.131	0.002	0.011
(standard error)	(0.003)	(0.004)	(0.003)	(0.001)	(0.003)	(0.001)	(0.002)
Observations	551	551	551	551	551	551	551
$\bar{X}_C$	0.432	0.399	0	0.007	0.145	0.002	0.015
(standard error)	(0.004)	(0.004)	(0)	(0.001)	(0.004)	(0.001)	(0.002)
Observations	560	560	560	560	560	560	560
Panel B: t test comparison of means of outcome variables for participants in groups B and C.							
$H_0 : \bar{X}_B = \bar{X}_C$ (p-values)	0.000***	0.002***	0.000***	0.033**	0.008***	0.770	0.121
Panel C: t test comparison of means of outcome variables for participants in groups A and C.							
$H_0 : \bar{X}_A = \bar{X}_C$ (p-values)	0.001***	0.000***	0.000***	0.121	0.000***	0.716	0.020**
Panel D: t test comparison of means of outcome variables for participants in groups A and B.							
$H_0 : \bar{X}_A = \bar{X}_B$ (p-values)	0.812	0.169	0.646	0.681	0.123	0.934	0.000***

Note: Seven use of time observations are missing for endline survey respondents. *, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively.

Chapter 3

Not compromising: the impact of a poverty graduation program on women's empowerment

3.1 Introduction

In 2015, the Sustainable Development Goals included the achievement of gender equality and empowerment of all women and girls as its fifth goal. This new goal reflects both the intrinsic value of gender equity and the instrumental value of increased female autonomy and its association with a variety of positive outcomes such as reduced fertility, increased child survival rates and the allocation of resources in favor of children within the household (see Duflo (2012) and the references therein).

These associations prompted many poverty alleviation programs to target women as their main beneficiaries and have women's empowerment as a key output (van den Bold et al., 2013). The commonly assumed causal pathway between such interventions and empowerment is that women who are provided with additional resources (such as transfers) will have greater bargaining power within the household. Such increase in bargaining power leads to an increased ability to make choices (Kabeer, 1999), from which the achievement of outcomes strongly aligned with their preferences (such as investments in the education or nutrition of children) then follow. Over time, gender relations and the social status of women within the household or the community may change (Duvendack, Palmer-Jones, & Vaessen, 2014).

Because changes in bargaining power, although central to the link between interventions and empowerment, are not directly measurable (J. Carter et al., 2014), most of the existing analysis relies on proxy measures such as self-reported participation in household decisions or control over assets, among others. Although these measures have often been criticized for their lack of rigour (van den Bold et al., 2013), measurement problems such as social desirability bias (Jejeebhoy,

2002), and for potential inadequacy to local context (Schatz & Williams, 2012; Upadhyay & Karasek, 2012) their use has been widespread in the evaluation of the impact of poverty graduation programs on women's empowerment. The general conclusion is that there is little evidence of an impact.

Banerjee, Duflo, et al. (2015), in an analysis of six poverty graduation programs across six countries, find that at the end of the program women have more say in decisions on health expenditures and home improvements but this gain does not persist after the end of the program, even though consumption and income gains do. They also show that the impact depends on the cultural context: the standardized mean treatment effect of poverty graduation programs on women's empowerment is larger in South Asian countries (India and Pakistan) compared to African (Ethiopia and Ghana) and Latin American countries (Peru and Honduras), which may either reflect initial differences in bargaining power or the inadequacy of the proxies used to measure empowerment. [1] Bandiera et al. (2016) similarly find no impact of a poverty graduation program, implemented in Bangladesh, on women's empowerment.

One alternative to survey-based measures, which we will follow, is to use incentivized decision making tasks that provide an opportunity to directly measure bargaining power. This is the approach used by de Palma, Picard, and Ziegelmeyer (2011), Carlsson, Martinsson, Qin, and Sutter (2013), Braaten and Martinsson (2015) and Almas, Armand, Attanasio, and Carneiro (2015). In the approach of de Palma et al. (2011), Carlsson et al. (2013) and Braaten and Martinsson (2015), relative bargaining power is determined by examining the influence of individual preferences on a couple's joint preferences, and inferred from the difference between decisions made individually and jointly by a husband and wife. Almas et al. (2015) take a different approach in their experiments and use women's willingness to pay to receive a transfer instead of their partners as a measure of empowerment. In contrast to these studies, the approach we implement reflects traditional concepts of power in the social sciences (Dahl, 1957; Harsanyi, 1962; Russell, 1938), and bargaining power is determined from a woman's decision to change her spouse's decision, hence providing a more direct measure of a woman's household bargaining power.

We use the decisions in this task to evaluate the impact on women's empowerment of a poverty graduation program targeted at ultra-poor women in northern Kenya, the Rural Entrepreneur Access Project (REAP), which is briefly presented in the next section. The program was rolled-out in three funding cycles, with beneficiaries being randomly allocated to each cycle. Survey measures of women's decision making power were collected at baseline and at two follow-up surveys.

[1] Many of the survey items that capture these measures are grounded in formative research from South Asia where women's empowerment may manifest differently (Malhotra, Schuler, & Boender, 2002; Schatz & Williams, 2012; Upadhyay & Karasek, 2012).

In addition, in mid-2014, the beneficiaries (and their spouses) from the first and last funding cycle, who benefited from the program one year apart from each other, were invited to a set of decision making experiments designed to directly measure women's bargaining power within the household. In the experiment, women (and their spouses) perform two tasks. Firstly, they make decisions on investments in a simple risk elicitation task based on Gneezy and Potters (1997). In a second task, they play a modified version of the first task in which the investment decision takes place in two stages. In the first stage, the first mover decides how much to invest in the risky lottery. The second mover then has the opportunity to change the decision. Because all participants first act as the first mover and then as the second mover in response to every possible first move by their spouse, an indicator of empowerment that reflects the woman's decision to not compromise from her own preferred investment can be defined. This indicator explicitly reflects traditional concepts of power (Dahl, 1957): the woman, in making a decision to change the husband's investment decision to one that her husband would not otherwise choose (but she would) is exercising her power over her husband. [2]

The two instruments (the survey questions and the lab-in-the-field design and procedures) are presented in detail in section 3.3. The evidence on women's empowerment is discussed in section 3.4. We find that the experimental measure has little correlation with survey measures. In section 3.5, estimates of the impact of REAP on these measures of empowerment are presented. Participation in REAP leads to an increases in empowerment when empowerment is measured using the experimental measure, but not when it is measured using the traditional survey measures. In section 3.6 both the experimental and survey measures are correlated with indicators of well-being that one would expect, based on the literature, to improve with women's power. The experimental measure is found to be a better proxy for empowerment compared to survey measures, based on its correlation with indicators of well-being. A conclusion is offered in section 3.7.

3.2 The Rural Entrepreneur Access Program (REAP)

The Rural Entrepreneur Access Program (REAP) was implemented in 14 locations in northern Kenya (see Figure 3.1), a region where more than 80% of the population are estimated to live below the national poverty line and where gender inequality is estimated to be more than 25% higher than the national average (Kenya National Bureau of Statistics and Society for International Development,

[2] Dahl (1957) offers the following definition of power: "A has power over B to the extent that he can get B to do something that B would not otherwise do."

2013). [3] In 2010, Kenya's new constitution explicitly gave women the same legal rights as men but despite improvements in legal gender equality, women in Kenya continue to suffer from gender inequities which are exacerbated in rural areas (Nature Conservancy, 2013). In northern Kenya, where pastoralism is the main livelihood option and livestock serves both as a source of income and food for herders and their families, social and cultural norms restrict women's ownership rights of livestock. Despite women playing a key role in the management of livestock, their contribution usually goes uncompensated and their ability to dispose of livestock is greatly limited, with men usually having the final say in this decision (Fratkin, 2004).

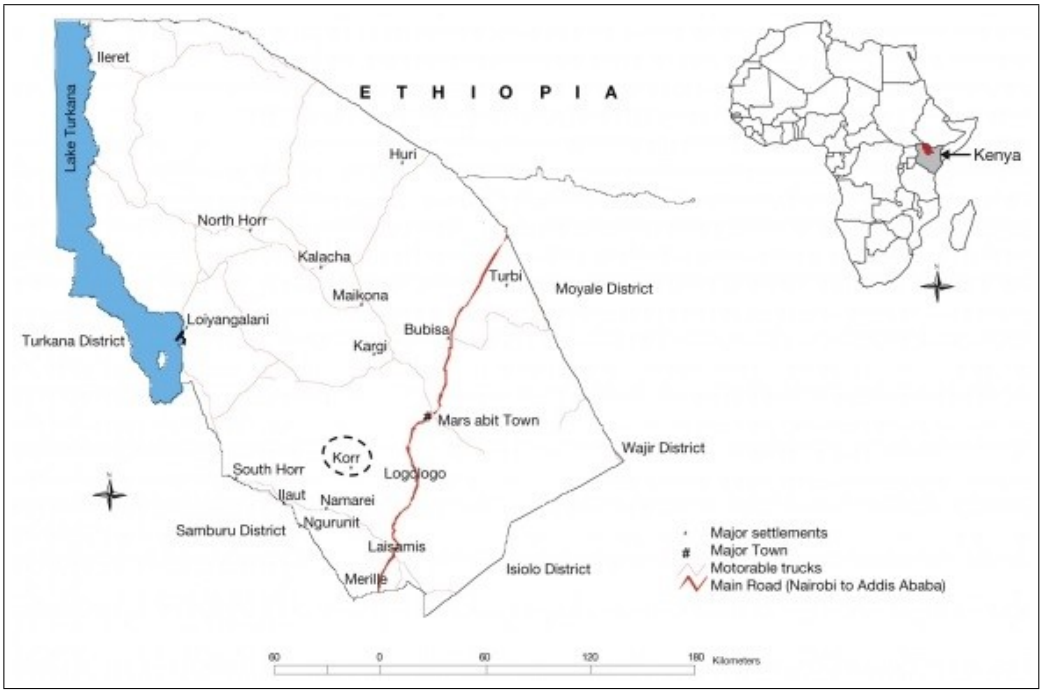


FIGURE 3.1: Map of Marsabit County (Warui & Kshatriya, 2009).

The program provides ultra-poor women in northern Kenya with multiple interventions, with the aim of giving them a localized "big push" to graduate from poverty by addressing the overlapping set of constraints that they face. This multifaceted approach to poverty, commonly referred to as the poverty graduation approach, was pioneered by BRAC through the program Challenging the Frontiers of Poverty Reduction – Targeting the Ultra-Poor (CFPR/TUP) (Goldberg & Salomon, 2011; Matin et al., 2008). During a limited period (2 years), participants of the CFPR/TUP program benefit from a set of interventions which includes initial consumption support and an asset transfer, together with savings services, skills training, and regular follow-up visits by BRAC program officers and livestock specialists, with the objective of improving returns from the asset

[3] In Marsabit County, where REAP was implemented, the Gender Inequality Index (GII) in 2013 was estimated to be 0.69 compared to the national average of 0.55 (Kenya National Bureau of Statistics and Society for International Development, 2013). The United Nations Human Development Report, 2015 ranks Kenya 126 out of 155 countries based on the GII.

transfer. The poverty graduation approach has been implemented and evaluated in many different contexts and has been found to result in sustained improvements in income, consumption, food security, but, as discussed previously, the evidence points to a limited impact on women's empowerment (Bandiera et al., 2016; Banerjee, Duflo, et al., 2015).

The Rural Entrepreneur Access Project (REAP) is similar in spirit to poverty graduation programs implemented elsewhere but also presents several notable differences. [4] First, REAP relies entirely on a cash transfer rather than a physical asset as a way to increase beneficiaries' wealth. Second, the program is explicitly enterprise focused, with the requirement that the women form three-person groups to jointly run the enterprise. Finally, REAP is implemented in the context of very limited market access, with an economy that is based on one activity (raising livestock) that is prone to frequent shocks due to drought. Such shocks force the poorest households to turn to alternative income generating activities, providing the opportunity for REAP to lead to the empowerment of women, as relatively poorer households become less dependent on (male dominated) livestock production for income and more dependent on other sources such as the enterprises established under REAP.

In November 2012, 1755 women were identified as being eligible for participation in REAP. Due to a lack of capacity to simultaneously enrol all eligible women, the set of potential beneficiaries was split into three groups to be successively enrolled over the next three funding cycles (March/April 2013, September/October 2013 or March/April 2014). Assignment to each cycle was done randomly, through a public lottery that took place in each of the locations from which participants had been recruited, with one-third of the women enrolled in each funding cycle. All eligible women were interviewed at baseline (November 2012) and at two follow-up surveys, conducted at six month intervals and timed to coincide with the beginning of each new funding cycle. Attrition in the follow-up surveys is very low with less than 2% of women not being reached for interviews in either the endline or midline surveys.

Gobin, Santos, and Toth (2016) take advantage of the sequential roll-out of the program, the randomized allocation to each funding cycle, the perfect compliance of observations to treatment and control groups, and the low attrition rate to identify the program impacts on income, expenditure, consumption, savings, livestock holdings, durable assets and food security. [5] They find that participation in REAP resulted in significant improvements in income, savings and asset

[4] The program is implemented through a NGO, The BOMA Project. See http://bomaproject.org/the-rural-entrepreneur-access-project/ for a complete description of REAP.

[5] Gobin, Santos, and Toth (2016) also check the validity of the experimental design. They find that the groups that were assigned to different funding cycles are balanced on baseline characteristics. They also show that spillover and program anticipation effects, if at all present, are of limited importance.

accumulation, and food security, with the increase in income and savings being driven by women's participation in microenterprises and savings groups. Such improvements in income earned by the woman may potentially strengthen her bargaining position within the household, a hypothesis that is tested in this study.

The data collected to estimate the impact of this program also includes women's reports of their participation in decisions about the purchase of food, household items, livestock, and paying for schooling and medical fees. This information is used to construct a household decision making power index (*HDMI*), a commonly used proxy indicator for empowerment, which is discussed in section 3.3.2. The alternative measures of empowerment come from several experiments for which a subset of beneficiaries of the program from the first and last funding cycle were invited. These experiments took place in June/July 2014, approximately 14 months (2 months) after the March/April 2013 (March/April 2014) funded groups began participating in REAP, and 2 months after the endline household surveys in the REAP impact evaluation study were conducted. [6] During these two months, participants in the March/April 2014 funding cycle started to benefit from REAP: they participated in a business skills training session and received a cash transfer to implement their business plan. In light of this, the impact of REAP on the experimental measure of empowerment estimated in this study will likely understate the true impact of REAP on empowerment. A detailed description of the experiment and experimental procedures is provided in section 3.3.1.

3.3 Measuring empowerment: surveys and experiments

3.3.1 Experimental design and procedure

In June 2014 married REAP beneficiaries who received funding in March/April 2013 and March/April 2014, along with their husbands, were invited to participate in a household decision making experiment. [7] Several steps were taken to maximize the attendance. Invitations were made by the business mentors, who were provided with a checklist of eligible beneficiaries to ensure that participants were either from the March/April 2013 or March/April 2014 funding cycle. [8] On the day before the experiment, mentors reminded all eligible persons of the experiment. The team of enumerators arrived in each village the day before the experiment, which also served as a reminder.

[6] See Appendix 3.A, Figure 3.A.1 for a timeline and sequence of activities in the three funding cycles, including the timing of experiments.

[7] The term married is used to refer to women who are either married or in a relationship.

[8] Business mentors are responsible for providing training and follow-up support to REAP beneficiaries in each location where the program is implemented.

The experiments were run separately in each location by a team of four enumerators and a research assistant who was in charge of overseeing all experimental procedures. [9] Three enumerators focused on conducting experiments with women and one focused on experiments with men. Two sets of tasks were conducted: the first set focused on household decision making and was run with both women and men, while the second set included an trust game and a coordination game, in addition to tasks designed to elicit risk and time preferences, and was run with women only. Women were only made aware of the second set of tasks after they had finished making their decisions in the first set, hence latter decisions are not expected to confound the decisions made beforehand. [10] An outline of the sequence of tasks that took place during the experimental sessions is presented in Table 3.1.

The experiment on household decision making takes the form of a risk elicitation task which, after being played first as an individual task, is modified to allow the decision of individuals to affect the outcomes of their spouses. The Gneezy and Potters (1997) risk elicitation task is used because it is easier to understand and implement than other approaches, which seemed of central importance in the setting studied, where beneficiaries have less than one year of formal education. [11] In this task, the participant receives an endowment of 100 Ksh and then decides how much to invest in a risky lottery. [12] The possible investments are restricted to 0, 25, 50, 75 or 100 Ksh. The investment is doubled with a probability of two-thirds and lost with probability one-third.

After the participant decides how much to invest in the first task, the enumerator proceeds to explain the second task. In this second task, the risk scenario is changed to a sequential move game between the participant and her spouse. The endowment is increased to 200 Ksh and participants and their spouses (who are physically separated) are each told that they will only receive half of the total payoff if this task is chosen for payment. The first mover decides how much of the 200 Ksh to invest in the risky lottery, with the investments restricted to 0, 50, 100, 150 or 200 Ksh. The lottery is otherwise identical to the lottery already played in the first (individual) task. The second mover then has the opportunity to accept

[9] Eight local language enumerators were trained for three days on the experimental procedures before being divided into two teams. Experiments took place in churches, schools or meeting halls which were divided into five separate spaces for each enumerator and the research assistant.

[10] If the woman was married and her spouse was present or she was uncertain if he would attend then the enumerator conducted both sets of experiments with her. If she knew that her spouse was not attending or she was single then the enumerator started with the second set of experiments.

[11] Other risk elicitation methods such as those used by Eckel and Grossman (2002) and Holt and Laury (2002) were also considered in a pilot study conducted in June 2013 but it was found that many women were making inconsistent choices when the Multiple Price List method was used, which was likely due to their lack of understanding of the task. See Charness, Gneezy, and Imas (2013) for a useful review of commonly used methods.

[12] At the time of the experiment the exchange rate was 1 USD to 85 Ksh and the average daily wage for menial labor was approximately 200 Ksh. The average daily consumption per capita for the study sample at endline was approximately 58 Ksh.

TABLE 3.1: Sequence of events at experimental sessions

Activity	Women				Men
	Single	Married			
		Husbands do not attend	Not sure if husbands will attend	Husbands attend	
Risk 1			✓	✓	✓
Household task			✓	✓	✓
Household decision making survey					✓
Risk 2	✓	✓	✓	✓	
Trust game	✓	✓	✓	✓	
Coordination game	✓	✓	✓	✓	
Payment	2 of 3	2 of 3	1 of first 2 and 2 of last 3 if husband attends **OR** Risk 1 and then 2 of last 3 if husband does not attend	1 of first 2 and 2 of last 3	1 of 2
Time preference: hypothetical	✓	✓	✓	✓	

Note: In this study we report on the items in bold print. The other tasks are part of the second set of experiments which are designed to look at decisions and preferences among business group members and are therefore not relevant to this study on household decision making.

the investment decision of the first mover or to change the decision. If the second mover changes the decision, s/he has to choose a new amount to invest.

The participant and their spouse each play the role of both the first mover and the second mover, which allows for the estimation of a measure of household decision making power based on the choices of the second mover relative to the initial investment of their spouse as well as their own preferred investment when they are the first mover. This experiment makes use of the strategy method in which a responder makes conditional decisions for each possible information set (i.e. 0, 50, 100, 150 and 200 Ksh). Such method is easier to implement in the field than the direct response method and also increases statistical power. Many studies that compare the two methods find no real difference between them in the case of sequential moves (see Brandts and Charness (2011) for a survey of this literature). [13]

Before beginning the experiments women naturally separated from men as they waited to enter the venue. [14] Once the participants consented to take part in the experiment, the rules of the first task were explained, first orally and then visually. [15] Two bowls were used to represent the money kept by the participant and the money invested in the lottery, with real money being used in the demonstration. The probabilities for the high and low payoffs in each part were demonstrated using four white and two blue balls. Several examples were used to further illustrate the task, before presenting the participant with a scenario to check their understanding. If the task was still not clear to the participant, the explanation was repeated until it was clear that the participant fully understood the task. The participants were informed that together, these two tasks would take approximately 20 minutes to complete and that at the end of the experiment they would (blindly) pick a numbered ball from a bag, to determine which task they would be rewarded for. The enumerators stressed that the payments would take place in private at the end of the sessions and that their decisions would not be revealed to other participants, including their spouses. There was no communication between spouses and participants were informed that a coin toss would determine if their first mover decisions or second mover decisions would be used in calculating their payoffs. Once a couple had completed both tasks they were invited separately to pick a numbered ball from a bag to determine which task from the first experiment would be played for real. If the first task was chosen

[13] However, it is argued that hot vs cold decisions might make a difference for experiments in which emotions are involved. For example, Brosig, Weimann, and Yang (2003) find higher punishment by second movers in response to a selfish play by the first mover when the direct response method was used. But Brandts and Charness (2011), in their review of the literature find no significant differences in direct response and strategy method for games in which emotions are involved although they do find that a particular aspect of emotions related behavior, the use of punishment, is more likely with direct response than strategy.

[14] The mentors also ensured that participants that were waiting to take part in the experiments were kept separate from those who had completed the experiments.

[15] No eligible participants who attended the sessions declined to participate in the experiments.

for payment then the participant picked a colored ball from a bag to determine if their investment was doubled or lost. If the second task was chosen, then a coin toss first identified whose first mover and second mover decisions would be used to determine payoffs. A colored ball was then picked to determine if the final amount invested was doubled or lost. If a woman was uncertain as to whether her husband would show up, she was asked to wait until he arrived. If he did not arrive then she only received payment for the first task (in addition to any other payment she received from the second experiment). The full set of instructions used by the enumerators is presented in Appendix 3.B.

By focusing on decision making under risk, our design is similar to several other experiments on household decision making. However, while much of the previous work typically elicits individual and joint preferences, and then examines the influence of individual preferences on joint preferences as a way to determine relative bargaining power within the household (for example, Braaten & Martinsson, 2015; Carlsson et al., 2013; de Palma et al., 2011), the tasks in this study were designed to directly capture classic concepts of power as defined, for example, by Dahl (1957) where person "A has power over B to the extent that he can get B to do something that B would not otherwise do". In this task, women can reveal their power by changing the investment decisions after the husband's preference is clearly stated.

We use the decisions made in this task to construct an indicator of decision making power in the following way. As a second mover, a woman must first decide to either accept or change the investment decision made by her spouse (I_1^H) when the husband's investment differs from her preferred investment (I_1^W). After that, and if a decision is made to change the husband's decision, the woman faces a second decision, of how much to invest instead (I_2^W). The use of the strategy method in this task results in five second mover decisions (I_2^W) for each woman, in response to the five possible first mover investments by her spouse, i.e. I_1^H= 0, 50, 100, 150 or 200 Ksh, although only four of these decisions are informative about intrahousehold differences in power (as we can learn nothing about these differences when $I_1^W = I_1^H$). [16]

By examining the frequency with which a woman changes her spouse's decisions one can gain some insight into household bargaining power. Does she simply choose her own preferred investment or does she compromise with the husband's decision? If she compromises with her spouse does she stick with her spouse's choice or choose some amount between her own preferred investment and her spouse's? For those cases when $I_1^W \neq I_1^H$, we can summarize these decisions

[16] A total of 700 women participate in the experiments resulting in 3500 second mover decisions being made. The sample of participants in the experiments is discussed further in section 3.4.

through a new variable, as follows:

$$compromise = \begin{cases} 1 & \text{if } I_2^W = I_1^H \\ 1 & \text{if } I_2^W \neq I_1^H \text{ and } I_2^W \neq I_1^W \\ 0 & \text{if } I_2^W \neq I_1^H \text{ and } I_2^W = I_1^W \end{cases} \tag{3.1}$$

Here, *compromise* is equal to zero if the woman decides to stick with her original investment decision, I_1^W, which differed from her husband's decision. Alternatively, the variable *compromise* is equal to one if she decides to change her husband's investment to some amount that is not equal to her original preference or if she chooses to accept her husband's decision. In addition to excluding those observations for which $I_1^W = I_1^H$ we also excluded inconsistent observations where, similarly, nothing stands to be learned from the decisions. For example those observations where $I_1^H > I_1^W > I_2^W$, $I_1^W > I_1^H > I_2^W$, $I_2^W > I_1^H > I_1^W$, or $I_2^W > I_1^W > I_1^H$ are excluded as these decisions involve the woman choosing an amount I_2^W that lies outside of the interval between her first mover investment, I_1^W, and her husband's first mover investment, I_1^H. The number of excluded observations for these reasons is relatively small (6.5% of the total number of useful observations, i.e., those left after the cases where $I_1^W = I_1^H$).

We use the multiple observations of *compromise* for each woman to construct a single index of each woman's *power*, which we define as a binary indicator taking the value of one if, more often than not, the wife does not *compromise* with her husband (i.e. the average is less than 0.5), and zero otherwise.

Although the scenario presented in this experiment is neutral it also reflects the many instances where women are charged with the responsibility of executing the household expenditure decisions, particularly those decisions regarding expenditures on food and other household items. This experimental scenario can be seen as capturing a woman's decision to go against the wishes of her husband regarding what items to purchase. In northern Kenya, it is not uncommon for husbands to rely on women to execute their wishes with regards to household expenditure when they migrate with livestock herds (or to towns). In many cases the woman is left with some livestock (or is sent money) and has the choice to either follow the husband's wishes or to go against them, a decision that is similar to the one faced in this experiment.

3.3.2 Survey measures

As part of the evaluation of the impact of the program, a baseline survey was conducted in November 2012, with all women selected to benefit from REAP being interviewed prior to the first group of women receiving funding. Follow-up

surveys were then conducted at six-month intervals to coincide with the roll-out of the program to beneficiaries in later funding cycles. [17] These surveys included a section on household decision making, with questions phrased as follows: "When you have to buy food, how is the decision made? Who has the final say?" Possible responses to this question were husband only, wife only, husband and wife, or other. In addition to food, questions were also asked about decisions over the purchase of livestock, the purchase of household items, and the payment of children's medical expenses and school fees.

Such questions are widely used as proxy indicators for empowerment with many studies combining the responses to these questions to construct a single index, where the response to each question is coded as one if the woman has the final say in the decision or zero if otherwise, and therefore identifies empowered women as those who have sole decision making powers over household decisions. The answers are then summed to form a household decision making index, which is labeled as *HDMI1*.

Recent work by Peterman, Schwab, Roy, Hidrobo, and Gilligan (2015) finds that substantially different conclusions on empowerment can be made depending on whether household decision making indicators include joint decision making as opposed to sole decision making only. Given this, the same responses are used to construct a second index of empowerment, labeled as *HDMI2*, in which the response to each question is coded as one if the woman has the sole or joint final say in the decision or zero if otherwise, and therefore identifies empowered women as those who have at least some say in household decisions.

3.4 Measuring empowerment: summary of experimental results

As mentioned in section 3.3.1, only married women and their spouses were eligible for these tasks, leaving us with a potential sample of 946 women who reported having a husband or being in a relationship (out of the 1167 women that participated in REAP in the two funding cycles that started one year apart). A total of 700 eligible women (and 33% of their spouses) accepted the invitation to participate in the experiment (see Table 3.2). [18]

We focus on the decisions made by women in the second task as these are the decisions used to derive our experimental measure of empowerment, *power*. As

[17] See Appendix 3.A, Figure 3.A.1 for the timing of surveys.

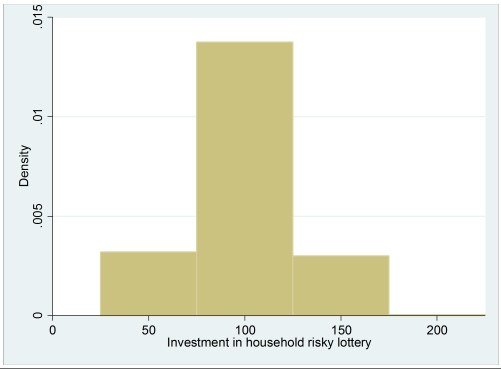
[18] Note that, although the experiment was designed for couples, they were not required to make decisions simultaneously. Women who were unsure whether their spouses would show up to take part in the experiment, were still allowed to participate. Summary statistics for the male respondents are presented in Appendix 3.C.

TABLE 3.2: Sample size

Funding Cycle:	Treatment Mar/Apr 2013	Control Mar/Apr 2014
Eligible REAP beneficiaries	585	585
Final REAP beneficiaries	585	582
Beneficiaries eligible for experiments	483	463
Beneficiaries that attend experiments	336	364
Husbands that attend experiments	118	111

the first mover in the second task the woman must decide how much of her endowment (of 200 Ksh) to invest in the lottery where her spouse has the chance to change her investment decision. Women's first mover decisions are summarised in Figure 3.2 where we show that on average they invest half of the endowment (99.786 Ksh).

FIGURE 3.2: Summary of investments by women in the household risky lottery

	Women
Mean	99.786
(Std Dev)	(28.367)
N	700
Distribution of investment decisions	

Note: The initial endowment given to participants to invest is 200 Ksh. Two women chose to invest the maximum of 200 Ksh.

The decision of women, as second movers, to change the spouse's decisions (for all possible first mover decisions where $I_1^W \neq I_1^H$ and excluding those observations where $I_1^H > I_1^W > I_2^W$, $I_1^W > I_1^H > I_2^W$, $I_2^W > I_1^H > I_1^W$, or $I_2^W > I_1^W > I_1^H$) is reported in Table 3.3. A greater proportion of women change the spouse's investment to their own preference when the spouse's decision is at an extreme (i.e. $I_1^H = 0$ or 200) and a greater proportion of women are less likely to change their spouse's decision if it lies between these two extremes, a result which is expected as women prefer investments that are away from the extremes (see Figure 3.2).

TABLE 3.3: Summary of women's decisions as second movers

I_1^H	$I_2^W = I_1^H$	$I_2^W \neq I_1^H$ and $I_2^W \neq I_1^W$	$I_2^W \neq I_1^H$ and $I_2^W = I_1^W$
0	0.233	0.165	0.602
50	0.691	0.037	0.272
100	0.873	0.000	0.127
150	0.586	0.079	0.336
200	0.230	0.244	0.527

Note: This table reports the proportion of women who, as second movers, choose to either accept the spouse's decision ($I_2^W = I_1^H$), change it to some amount not equal to her original preference ($I_2^W \neq I_1^H$ and $I_2^W \neq I_1^W$) or stick with her original preference ($I_2^W \neq I_1^H$ and $I_2^W = I_1^W$), for each possible first move decision by the spouse (I_1^H). The sample used here excludes those observations where $I_1^W = I_1^H$, $I_1^H > I_1^W > I_2^W$, $I_1^W > I_1^H > I_2^W$, $I_2^W > I_1^H > I_1^W$, or $I_2^W > I_1^W > I_1^H$.

Summary statistics for the *compromise* and *power* variables for women are presented in Table 3.4. On average, women chose to change their spouse's investment to their own preferred investment in more than 40% of the cases, and more than 25% of women change more than half of the spouse's possible first mover decisions to their own preferred investment (i.e., do not compromise in more than half of the cases).

TABLE 3.4: Summary of the experimental measure *power*, and its constructs.

	compromise	*compromise* averaged across first mover decisions	*power*
Mean	0.583	0.598	0.266
(Std Dev)	(0.493)	(0.313)	(0.442)
N	2619	700	700

Note: The variable *compromise* is a binary indicator equal to zero if the woman decides to stick with her original investment decision which differs from her spouse's decision. The variable is equal to one if she changes her spouse's decision to some amount not equal to her original investment or if she chooses to accept her spouse's decision.The average of *compromise* varies from zero (if the woman never compromises) to one (if she always compromises). The variable *power* is coded as one if in more than 50% of decisions as the second mover, the woman does not compromise at all with her husband.

Before presenting estimates of the impact of REAP on bargaining power, we present the correlation between experimental and survey measures of empowerment. The correlation coefficients are reported in Table 3.5. We find a high correlation between *HDMI1* and *HDMI2*, which was to be expected given the way these indexes are constructed. However, *power* is not correlated with either of the survey measures of empowerment, a conclusion that is similar to Almas et al. (2015) who also use an incentivized task to derive a quantitative measure of empowerment, and similarly find little correlation between their experimental

measure and individual survey measures. [19] It is important to notice that, were the two measures largely correlated, it is likely that preference should be given to measuring empowerment using survey measures, given that their implementation is both easier and less expensive.

TABLE 3.5: Correlations between different measures of empowerment.

	HDMI1	*HDMI2*	*compromise* (average)	*power*
HDMI1	1.000			
HDMI2	0.557 (0.000)	1.000		
compromise (average)	-0.025 (0.511)	-0.016 (0.672)	1.000	
power	0.039 (0.313)	0.012 (0.748)	-0.815 (0.000)	1.000

Note: Parentheses contain p-values of the test of the null hypothesis that the two variables are independent. The components of *HDMI1* are coded as one if the woman has the final say in the decision or zero if otherwise whereas those of *HDMI2* are coded as one if the woman has the sole or joint final say in the decision or zero if otherwise. *HDMI1* and *HDMI2* are standardised using the control group mean and standard deviation. The variable *compromise* (average) is the average of *compromise* across all possible first move decisions of the spouse. The variable *power* is coded as one if in more than 50% of decisions as the second mover, the woman does not *compromise* at all with her husband.

There are three explanations for this result. The first is that both experimental and survey measures capture substantially different dimensions of empowerment. The second is that one of the indexes does not appropriately measure bargaining power, either because the experiment is too alien to women's decisions to meaningfully capture such dimension, or because that is true for survey measures, developed and calibrated in one specific context that is substantially different from the one we study (Schatz & Williams, 2012; Upadhyay & Karasek, 2012). [20] The third is that since the design of the game makes it difficult for a husband to know his wife's decision, her decisions in the game may be poorly correlated with real world decisions. We address these alternative explanations in two ways. The first, discussed in the next section, by quantifying the impact of REAP on the different measures of bargaining power: if we believe in the causal pathway from assets to agency that generally underlies the discussion of empowerment, then the identification of an impact of REAP on one of the measures of power would go someway into validating that measure. The second way, similar

[19] Almas et al. (2015) do find that their experimental measure of empowerment is correlated with an index of the survey measures but this correlation goes against the expected direction.

[20] For example, the module on women's empowerment in the Demographic and Health Surveys (which contains questions similar to those used in this study) have been criticized for not being adequate for use in Sub-Saharan Africa (Schatz & Williams, 2012; Upadhyay & Karasek, 2012) with concerns arising because the survey items are grounded in formative research from South Asia (Malhotra et al., 2002) where women's empowerment may manifest differently.

in spirit and discussed in section 3.6, is to estimate the correlation between the measures of power and welfare indicators that are usually taken to measure the benefits of empowerment.

3.5 The impact of REAP on empowerment

The following regression specification is used to estimate the impact of REAP on the measures of empowerment (both survey-based and experimental) described above:

$$Y_i = \theta + \beta T_i + \varphi X_i + \epsilon_i \tag{3.2}$$

where Y_i is the outcome of interest for woman i, which is either *HDMI1*, *HDMI2* or *power*, T_i is a binary variable equal to one if woman i benefited from REAP in March/April 2013 and zero if she benefited in March/April 2014 (hereafter referred to as treatment and control groups respectively), and X_i is a matrix of control variables that includes sub-location fixed effects as well as controls for the woman (as well as her spouse's age), her literacy and numeracy, whether she is in a polygamous marriage, the number of children and adults in the household, and the livestock and other durable assets owned by the household, all at baseline levels. [21] When Y_i is *power*, control variables for the husband's attendance at the experiments as well as the woman's investment decision in the individual risk task are also included.

To interpret the estimates of β in a causal way, we need to address the possible bias associated with the fact that not all eligible REAP beneficiaries participated in the decision making experiments: as mentioned above, a total of 700 of the 946 women that were eligible to participate in the experiments on household decision making accepted the invitation to attend the sessions. As a first step in addressing concerns about potential self-selection bias, and how it may bias the estimates of the impact of REAP on women's empowerment, we check whether, conditional on being eligible, participants in the experiments are different from non-participants in terms of baseline characteristics (collected in November 2012, prior to the first group of women being enrolled in REAP). The results of that comparison are presented in Table 3.6. Panel A presents summary statistics (means and standard errors) of both household and individual characteristics of participants and non-participants, while panel B presents the t-tests of the null hypothesis of equality at baseline.

It is clear from the analysis of this table that participants and non-participants are similar along many dimensions but that some notable differences do exist.

[21] Stratification took place at the sub-location level (77 sub-locations).

TABLE 3.6: Summary statistics and baseline comparisons between participants and non-participants in the experiment.

Variable:	Monthly income per capita	Monthly expenditure per capita	Monthly food expenditure per capita	Monthly non-food expenditure per capita	Total savings per capita	TLU per capita	Durable asset index	Meals per day	Nights that child has gone to bed hungry in past week	Proportion of children in school	Household Size	# children	Years of education	Business Experience	Benefitting from HSNP	Participating in CARE VSLA	HDMI1	Funding cycle
Panel A: Means and standard errors of variables at baseline																		
Participated	32.919	23.781	9.137	21.468	3.705	0.655	-0.153	1.948	0.580	0.419	5.899	3.887	0.379	0.556	0.130	0.090	-0.052	0.480
(Std Err)	(1.056)	(0.845)	(0.501)	(0.781)	(0.316)	(0.027)	(0.156)	(0.014)	(0.026)	(0.010)	(0.070)	(0.065)	(0.060)	(0.019)	(0.013)	(0.011)	(0.038)	(0.019)
Observations	700	700	700	700	700	700	700	694	693	698	700	700	700	700	700	700	684	700
Did not participate	31.073	20.052	11.021	19.430	4.405	0.667	0.300	1.894	0.534	0.391	5.797	3.813	0.252	0.500	0.081	0.098	0.148	0.598
(Std Err)	(1.803)	(1.027)	(1.308)	(1.201)	(0.987)	(0.045)	(0.278)	(0.027)	(0.040)	(0.019)	(0.111)	(0.106)	(0.084)	(0.032)	(0.017)	(0.019)	(0.063)	(0.031)
Observations	246	246	246	246	246	246	246	246	242	245	246	246	246	246	246	246	242	246
Panel B: t test comparison of means of baseline characteristics based on participation in experiments																		
p-value	0.377	0.005***	0.180	0.156	0.500	0.807	0.156	0.075*	0.339	0.177	0.437	0.551	0.222	0.134	0.025**	0.729	0.007***	0.001***
Panel C: F-test from regression of participation in experiments on variables above.[a]																		
	F-Stat	p-value																
	2.65	0.000***																

Note: All monetary values are reported in 2014 USD, PPP terms. TLU refers to tropical livestock unit which is a standardised way of measuring the size of a mixed herd: 1 head of cattle is equivalent to 0.7 camels, 10 sheep/goats, or 2 donkeys. A description of the durable asset index can be found in Gobin, Santos, and Toth (2016). The components of *HDMI1* are coded as one if the woman has the final say in the decision or zero if otherwise. *HDMI1* is standardised using the mean and the standard deviation of the overall sample of married women at baseline. Funding cycle equals one (zero) if assigned to Mar/Apr 2013 (2014). [a]Monthly food and non-food expenditure per capita are excluded from this regression.*, ** and ** stand for significant at the 10%, 5% and 1% level of significance, respectively.

Non-participant households spent less on food per capita, are less likely to have benefited from cash transfer programs targeted at the poor (the Hunger Safety Net Program (HSNP)), reported having more say in household decisions, and are more likely to belong to the treatment group (March/April 2013). Overall, it seems that participant households are observationally different from non-participant households, a finding that is reinforced by the results of a F-test of the joint effect of these variables on participation in the experiments, reported in panel C.

To address the potential bias from the self-selection into the experiments, observations are reweighted such that the distribution of observable pre-treatment characteristics is identical for participants and non-participants (Nichols, 2007). To do this, the relation between participation in experiments on observable pre-treatment characteristics, namely baseline levels of the household decision making index (a measure of pre-treatment female empowerment), income, expenditure, savings and assets, food security indicators, household composition, characteristics of the woman and her spouse (age, literacy, business experience, whether in a polygamous marriage), participation in other NGO programs, and cycle of assignment to REAP, is estimated. The estimates of the logit model used in explaining the participation decision are reported in Appendix 3.D (Table 3.D.1). [22] The estimates of the conditional probability of being a non-participant (i.e. the propensity score), $\hat{p}$, and the non-participation odds, $\hat{p}(1-\hat{p})$, are then used to reweight the participant observations to estimate the average treatment effect on the treated (Nichols, 2008).

3.5.1 Estimates of the impact of REAP on empowerment

Table 3.7 presents estimates of the impact of REAP on *power*, the experimental measure of empowerment as described above, both without (columns (1) and (3)) and with (columns (2) and (4)) additional controls. In columns (3) and (4) we present the reweighted estimates. [23] Our preferred estimates, in column (4), show that women who received funding from REAP one year earlier are more than 7%

[22] The –pscore– command in Stata is used to estimate the propensity scores ensuring that the balancing property is met. The analysis of the balancing property is restricted to participants and non-participants in the region of common support. Twenty-three participant observations are lost due to either missing values or their propensity scores lying outside the region of common support. The final sample comprises of 349 treatment and 328 control individuals.

[23] Simple t test comparisons and Mann-Whitney tests of differences in the measures of empowerment between treatment and control participants are presented in Appendix 3.E. There are no significant differences between treatment and control participants across any of the measures of empowerment but it should be noted that the observations used in these tests are not reweighted to account for self-selection into the experiment.

more likely to not compromise with their husbands when making decisions as the second mover. [24]

TABLE 3.7: The impact of REAP on *power*.

	(1)	(2)	(3)	(4)
Treatment	0.058*	0.071**	0.053	0.071*
	(0.034)	(0.035)	(0.036)	(0.037)
Age		-0.001		0.001
		(0.003)		(0.003)
Age of spouse		0.001		-0.001
		(0.002)		(0.003)
Literacy		-0.032		0.079
		(0.088)		(0.130)
Numeracy		-0.024		-0.058
		(0.066)		(0.069)
Wife in polygamous marriage		0.006		-0.041
		(0.039)		(0.040)
# children in household		0.002		-0.002
		(0.011)		(0.012)
# adults in household		-0.029		-0.059*
		(0.031)		(0.034)
TLU per capita		0.033		0.056*
		(0.031)		(0.031)
Durable asset index		0.002		0.003
		(0.005)		(0.006)
Husband attends experiment		0.020		0.003
		(0.043)		(0.045)
Investment in individual risk task		-0.001		-0.000
		(0.001)		(0.001)
N	700	687	677	665
R-squared	0.137	0.154	0.120	0.143

Note: Results from a linear probability model. The dependent variable *power* is coded as one if in more than 50% of decisions as the second mover, the woman does not compromise at all with her husband. Regressions include sub-location fixed effects. Demographic and household controls are at baseline levels. Regressions in columns (1) and (2) are unweighted. Regressions in columns (3) and (4) include the weights generated from the propensity score estimates. Standard errors are shown in parentheses. *, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

Participation in REAP has a positive and statistically significant impact on empowerment, with *power* being 28% (0.16 SDs) higher for participants who received funding earlier compared to those who received funding later. This difference is significant at the 6% level but the effect is not robust to the exclusion of control variables, in which case it is only significant at the 15% level. But this impact is

[24] In Appendix 3.F results are shown for the impact of REAP on the average of *compromise* across all first mover decisions, the variable that is used to construct *power*. Participation in REAP does not have a significant impact on this measure.

likely to underestimate the true impact of REAP on empowerment since at the time of the experiments all participants had received funding. [25] One might expect changes in empowerment as a result of participation in REAP to be a gradual process as women grow their enterprises, incomes and agency. However, we cannot dismiss the fact that changes in empowerment might take place quite early on in the program especially given the findings of Gobin, Santos, and Toth (2016) who show that after only six months of participation in REAP beneficiaries' incomes are significantly higher, with this effect being driven by income earned from the REAP enterprise. In addition, it is important to notice that estimate only measures the impact of the program when women are at the half-way mark of their participation in REAP (i.e. one of two years); it is possible that impact is would be larger for women who benefit from the program in its entirety.

The impact of REAP on survey measures of empowerment is reported in Table 3.8. Turning first to the impact of REAP on *HDMI1*, the index that reflects power of the woman as the sole decision making power, we find that this measure of empowerment increases by 0.007 SDs when the treatment group is reweighted using propensity scores and all control variables (apart from sub-location fixed effects) are excluded (column (5)). When all control variables are included, the effect is of the opposite sign, with participation in REAP resulting in a 0.014 SDs *decrease* in *HDMI1* (column (7)). However, these changes are small relative to the control group and are not statistically significant at conventional levels.

Turning next to *HDMI2*, which reflects power of the woman if she has at least some say in household decisions (i.e. has joint decision making powers), the estimated impact is consistently negative across the reweighted specifications (columns (6) and (8)), and, as with *HDMI1*, they are relatively small compared to the control group, and not statistically significant. Taken together, these results suggest that participation in REAP does not result in the empowerment of women as measured by either specification of the decision making index. [26]

There are several possible explanations for the fact that the estimates of the impact of REAP on the survey measures of empowerment are not statistically significant. It can be that the time horizon might be too short to expect to see changes in women's role in decisions over the various items that comprise these indexes. Alternatively, responses in the survey may be subject to social desirability bias or these survey measures may not capture empowerment in this context. The

[25] As noted previously, when the experiments took place the treatment group had been enrolled in REAP for approximately 14 months and the control group for 2 months.

[26] As a robustness check, the impact of REAP on these survey measures is also estimated for the entire sample of married women i.e. including those who did not participate in the experiment (but for whom survey measures were collected during the endline survey), and for which concerns about possible selection bias are not relevant. The results from these robustness checks are reported in columns 1 to 4 of Table 3.8, and confirm that participation in REAP does not lead to any significant change in empowerment, when survey-based measures are used to quantify it.

TABLE 3.8: The impact of REAP on survey measures of household decision making

Dependent variable:	(1) HDMI1	(2) HDMI2	(3) HDMI1	(4) HDMI2	(5) HDMI1	(6) HDMI2	(7) HDMI1	(8) HDMI2
Treatment	0.026	-0.012	0.048	0.033	0.007	-0.059	-0.014	-0.052
	(0.067)	(0.065)	(0.067)	(0.065)	(0.076)	(0.077)	(0.076)	(0.077)
Age			0.005	0.012**			-0.005	0.000
			(0.005)	(0.005)			(0.006)	(0.006)
Age of spouse			0.010*	-0.006			0.012**	-0.007
			(0.005)	(0.005)			(0.005)	(0.006)
Literacy			0.094	0.022			-0.031	-0.144
			(0.178)	(0.173)			(0.268)	(0.272)
Numeracy			0.324**	-0.025			0.384***	0.060
			(0.125)	(0.122)			(0.143)	(0.144)
Wife in polygamous marriage			-0.052	0.205***			-0.000	0.258***
			(0.075)	(0.073)			(0.082)	(0.083)
# children in household			-0.048**	-0.042**			-0.030	-0.045*
			(0.022)	(0.021)			(0.025)	(0.026)
# adults in household			-0.280***	-0.175***			-0.247***	-0.106
			(0.060)	(0.059)			(0.071)	(0.072)
TLU per capita			-0.040	-0.116**			-0.027	-0.074
			(0.060)	(0.058)			(0.064)	(0.065)
Durable asset index			0.014	0.021*			-0.001	0.006
			(0.012)	(0.012)			(0.011)	(0.012)
N	888	888	875	875	665	665	665	665
R-squared	0.146	0.223	0.189	0.250	0.146	0.153	0.248	0.288

Note: Regressions include sub-location fixed effects. Demographic and household controls are at baseline levels. Regressions in columns 1 to 4 use the sample of all married women enrolled in REAP in Mar/Apr 2013 and Mar/Apr 2014 regardless of participation in the experiment. Regressions in columns 5 to 8 only look at married women who participated in the experiments and also include the weights generated from the propensity score estimates. The components of the dependent variable *HDMI1* are coded as one if the woman has the final say in the decision or zero if otherwise whereas those of *HDMI2* are coded as one if the woman has the sole or joint final say in the decision or zero if otherwise. *HDMI1* and *HDMI2* are standardised using the control group mean and standard deviation, allowing the estimate to be interpreted as the effect size relative to the control group. Standard errors are shown in parentheses. *, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

latter possibility is examined in more detail in the next section 3.6 where the experimental and survey measures of empowerment are correlated with household outcomes that are expected to be higher for more empowered women.

3.6 The relationship between measures of empowerment and welfare outcomes

The emphasis on empowering women as a development strategy reflects its perceived association with several positive outcomes, including increased food security and the allocation of resources in favor of children. If this expectation is valid, we should be able to conclude, that higher levels of empowerment are associated with better household outcomes, specifically, in our sample, monthly food consumption per capita and the number of nights that a child has gone to bed hungry in the past week. [27] We use this assumption as a test of the validity of the two types of empowerment indicators. The OLS estimates of these correlations,

[27] However, this association between empowerment and household outcomes would be dampened if spouses in our sample had identical preferences for these outcomes (see, for example, Kusago and Barham (2001)).

after controlling for other variables that might be associated with the indicators of well-being (including woman's age (as well as her spouse's age), her literacy and numeracy, whether she is in a polygamous marriage, the number of children and adults in the household, and the livestock and other durable assets owned by the household) are presented in Table 3.9. [28]

TABLE 3.9: The relationship between measures of food security/consumption and measures of empowerment

Dependent variable:	# nights that child has gone to bed hungry in past week			Monthly food consumption per capita		
	HDMI1	*HDMI2*	*power*	*HDMI1*	*HDMI2*	*power*
Measure of empowerment	0.038	0.282***	0.088	1.040	1.725**	7.429***
	(0.052)	(0.050)	(0.108)	(0.869)	(0.857)	(1.776)
Age	0.024***	0.023***	0.023***	0.129	0.123	0.114
	(0.008)	(0.008)	(0.008)	(0.134)	(0.134)	(0.133)
Age of spouse	-0.017**	-0.014**	-0.016**	-0.421***	-0.396***	-0.404***
	(0.007)	(0.007)	(0.007)	(0.116)	(0.115)	(0.114)
Literacy	-0.252	-0.196	-0.263	13.776**	13.994**	13.141**
	(0.348)	(0.340)	(0.348)	(5.638)	(5.626)	(5.564)
Numeracy	0.092	0.090	0.110	-25.651***	-25.345***	-24.884***
	(0.179)	(0.174)	(0.178)	(3.008)	(2.983)	(2.950)
Wife in polygamous marriage	0.267**	0.190*	0.271***	2.149	1.703	2.470
	(0.103)	(0.102)	(0.103)	(1.728)	(1.738)	(1.706)
# children in household	-0.008	0.004	-0.009	-7.331***	-7.285***	-7.353***
	(0.032)	(0.031)	(0.032)	(0.536)	(0.535)	(0.528)
# adults in household	0.142	0.160*	0.140	-3.899**	-3.971***	-3.721**
	(0.092)	(0.089)	(0.092)	(1.509)	(1.493)	(1.477)
TLU per capita	-0.012	0.009	-0.018	1.517	1.611	1.118
	(0.080)	(0.078)	(0.080)	(1.334)	(1.333)	(1.319)
Durable asset index	0.000	-0.001	-0.000	0.481**	0.469*	0.459*
	(0.014)	(0.014)	(0.014)	(0.239)	(0.238)	(0.236)
N	654	654	654	665	665	665
R-squared	0.153	0.196	0.153	0.613	0.615	0.623

Note: The dependent variable is either the number of nights that a child is reported as going to bed hungry in the past week, or monthly food consumption per capita which is measure in USD PPP at 2014 prices. All regressions include sub-location fixed effects and are reweighted using the weights generated from propensity scores. Demographic and household controls are at baseline levels. The components of the *HDMI1* are coded as one if the woman has the final say in the decision or zero if otherwise whereas those of *HDMI2* are coded as one if the woman has the sole or joint final say in the decision or zero if otherwise. *HDMI1* and *HDMI2* are standardised using the control group mean and standard deviation. The variable *power* is coded as one if in more than 50% of decisions as the second mover, the woman does not compromise at all with her husband. Standard errors are shown in parentheses. *, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

As with other studies, our results are somewhat mixed. [29] Although both the survey and experimental measures of empowerment are positively correlated with the number of nights that a child goes to bed hungry in the past week, a result

[28] The correlations reported in Table 3.9 are robust to the exclusion of the additional control variables.

[29] For example, Peterman et al. (2015) compare two indexes that are similar to *HDMI1* and *HDMI2*, to household consumption, and find that for a sample of women from Ecuador, there are significant correlations between the decision making indexes and household consumption but no associations between the indexes and dietary diversity. They find the opposite in Yemen, with no associations between the indexes and household consumption but significant correlations between the indexes and dietary diversity.

that goes against the expectation that more empowered women would be able to reduce the incidence of children's hunger, this correlation is only statistically significant at conventional levels for *HDMI2*, suggesting that this indicator may be a poor proxy for empowerment in this context. However, there is a positive correlation between *HDMI2* and monthly food consumption per capita suggesting that households with more empowered women have higher levels of consumption. There is also a positive correlation between *power* and monthly food consumption per capita which is statistically significant at the 1% level. Taken as a whole, these associations suggest that *power* as measured by the experiment, is a better indicator of empowerment than either *HDMI1* or *HDMI2* in this context.

3.7 Conclusion

This study presents an incentivized decision making investment task specifically designed to elicit an objective measure of intra-household bargaining power. In this task, a woman is allowed to change her husband's investment decision to a decision that she favors. Such decision is closely aligned with classic conceptualizations of power such as Dahl (1957), with more powerful women choosing their own preferred investment levels, even when the husband's preference is clearly stated. Despite this advantage, experimental measures are more difficult to implement and more expensive than traditional survey based questions. Hence, it is important to understand whether there is, practically, an advantage in using the former over the latter.

We contrast the two measures in three steps. The first is a simple analysis of the correlation between the different indicators of power, both survey-based and experimental. We conclude that they are not correlated. There are two interpretations of this lack of correlation: either one measure is an adequate measure of power and the other is not, or they measure different (and orthogonal) components of women's agency. In the next two steps, we explore the evidence in favor of one of these alternative explanations in two ways. The first relies on the conceptualized causal pathway from assets to agency that underlies much of the expectation of positive effects of programs such as microfinance, cash transfers or poverty graduation programs on empowerment.

We use data from a poverty graduation program, the Rural Entrepreneur Access Project (REAP), to quantify the impact of the set of interventions under this program on empowerment. We conclude that the program has a positive and significant effect on empowerment, if we capture it using the experimental measure that we develop, but that we would not be able to capture this impact had we limited ourselves to quantify empowerment using traditional survey measures.

The second way relies on the possible extrapolation of the behavior in the experiment to other decisions outside this environment. If we are correct in the interpretation of the decision making process during the experiment, it follows that women who demonstrate more power in this task may also be more likely to bargain with their husbands when it comes to decisions where their preferences may not be aligned. We estimate the correlation between this measure and household outcomes to check the validity of this assumption and conclude that it is a better indicator of empowerment than either *HDMI1* or *HDMI2* in this context.

Additionally, concerns have been raised about the adequacy of survey items to capture empowerment in different cultural contexts since they were grounded in formative research from South Asia (Malhotra et al., 2002). This inadequacy might explain the discrepancies found by Banerjee, Duflo, et al. (2015), where poverty graduation programs, similar to REAP, were found to have a larger impact on women's empowerment in Asian countries (India and Pakistan) compared to African (Ethiopia and Ghana) and Latin American (Peru and Honduras) countries. With this is mind, it is argued that the neutrality of the experiment allows for the reliable estimation of intra-household bargaining power across different contexts, but this remains to be tested.

Appendix

3.A Timeline of activities

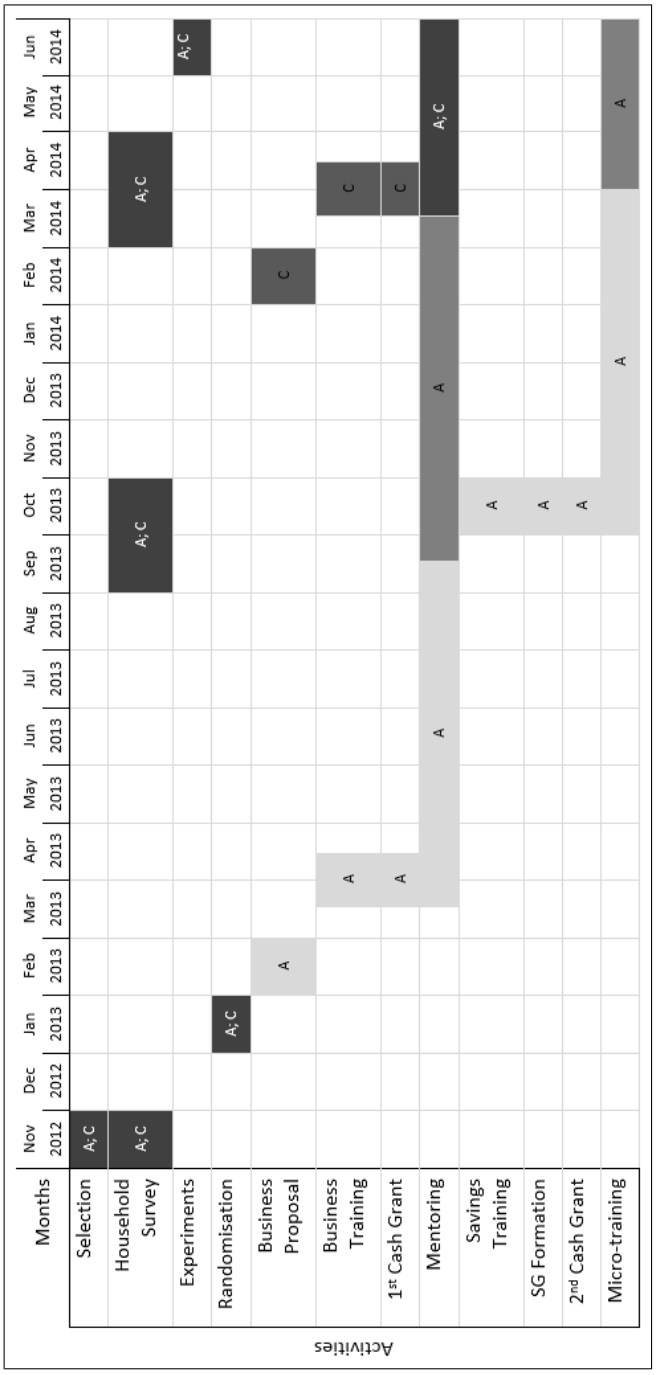


FIGURE 3.A.1: Timeline of data collection and program activities, by assignment to funding cycle.

Note: We label the beneficiaries who entered the program in the Mar/Apr 2013 funding cycle as group *A*, and those who entered in Mar/Apr 2014 as group *C*.

3.B Instructions for household experiment

Instructions

This section of the study will take approximately 20 minutes. There are 2 parts to this section and each will be explained at the appropriate time. Your earnings for Parts 1 and 2, and your total earnings for the study will be determined by the decisions you and the other Players make in each part.

You are free to make as much money as you can. How your rewards for Parts 1 and 2 will be determined is explained below. At the end of this section you will be presented with a bag with balls numbered 1 and 2. You will be asked to select ONE ball and the number on this ball will correspond to the Part of the section that you will receive money for. For example, if you pick balls numbered 1 then you will be rewarded for Part 1.

You will be paid in cash in private at the end of the session.

The decisions you make and the amount of money you receive will be held in strict confidentiality and will not be revealed to any of the other participants.

Part 1

In this part of the study you will work individually. You will be given 100 Ksh and will be asked to make an investment decision. You may choose to invest 0 Ksh, 25 Ksh, 50 Ksh, 75 Ksh or 100 Ksh.

If you choose to invest 0 Ksh then you will just keep the 100 Ksh you are given. If you choose to invest any amount greater than zero then you will be presented with an opportunity to double your investment. How?...

This bag is filled with 6 balls: 4 WHITE and 2 BLUE. You will be given a SINGLE chance to pick ONE ball (without looking) and the returns to the investment are based on the colour of the ball that is blindly selected from a bag. There are two possible outcomes of this investment: 1) If you pick a WHITE ball then the money you invested will be doubled and returned to you. 2) If you pick a BLUE ball then you will lose your investment.

Note, you will always retain the amount that is NOT invested regardless of the colour of the ball that is picked from the bag.

Example 1:

Suppose you invest 50 Ksh and keep 50 Ksh. You pick a WHITE ball. Your investment will automatically double to 100 KSH and you will receive a total of 150 Ksh i.e. 100 Ksh plus the 50 Ksh that was not invested.

Example 2:

Suppose you invest 50 Ksh and keep 50 Ksh. You pick a BLUE ball. Your investment will be lost and you will receive a total of 50 Ksh i.e. the amount that was not invested.

Question:

How much money would you receive in total if you invested 75 Ksh and you pick a WHITE ball?

Remember that in this task the bag will be filled with 4 WHITE balls and 2 BLUE balls

How much money do you choose to invest?

You will pick a ball from the bag at the end of the session if this task is selected for payment.

Part 2

In this task you will be paired with your spouse/partner and any rewards will be divided in half and you will receive half of the overall reward.

You will be asked to make an investment decision like you did in the previous task, i.e. whatever you invest is doubled if a WHITE ball is picked but the investment is lost if a BLUE ball is picked. The bag will contain 4 WHITE balls and 2 BLUE balls.

This time the amount you and your spouse/partner are given to invest is 200 Ksh.

You and your spouse/partner will be assigned as either "first mover" or "second mover" and this will be determined by flipping a coin with "heads" resulting in you being assigned as "first mover" and your spouse/partner being nominated as "second mover". If the coin lands on "tails" then the assignment will be reversed.

The game will be played as follows:

- The first mover will choose an amount to invest. They can choose to invest 0 Ksh, 50 Ksh, 100 Ksh, 150 Ksh or 200 Ksh.

- The second mover will then have a chance to accept or overrule the investment decision made by the first mover.

- If the second mover chooses to accept the investment decision of the first mover then a ball is chosen to determine if the investment is doubled or lost.

- If the second mover chooses to overrule the decision then they will be asked to select an alternative amount to invest. A ball will then be picked to determine if this new investment is doubled or lost.

- The reward for you will be half of the amount that is not invested plus half of the returns from the investment.

There will be no communication between the first and second mover.

Example:

The first mover chooses to invest 50 Ksh. The second mover then decides to overrule this decision and invest 150 Ksh instead. A WHITE ball is picked. The 150 Ksh investment is doubled and the total reward, including the amount that is not invested, is 350 Ksh which is divided in half.

The first mover chooses to invest 50 Ksh. The second mover then decides to accept this decision. A WHITE ball is picked. The 50 Ksh investment is doubled and the total reward, including the amount that is not invested, is 250 Ksh which is divided in half.

Question:

The first mover decides to invest 100 Ksh. The second mover decides to accept this decision. What is the total earned by the pair if a BLUE ball is chosen?

The first mover decides to invest 100 Ksh. The second mover decides to overrule this decision and invest 50 Ksh instead. What is the total earned by the pair if a BLUE ball is chosen?

Remember, there are 4 WHITE balls and 2 BLUE balls in the bag.

Suppose you are the first mover. How much will you choose to invest?

Suppose you are the second mover would you accept or overrule the first mover if they choose to invest 0 Ksh?

If overrule then how much would you invest?

Suppose you are the second mover would you accept or overrule the first mover if they choose to invest 50 Ksh?

If overrule then how much would you invest?

Suppose you are the second mover would you accept or overrule the first mover if they choose to invest 100 Ksh?

If overrule then how much would you invest?

Suppose you are the second mover would you accept or overrule the first mover if they choose to invest 150 Ksh?

If overrule then how much would you invest?

Suppose you are the second mover would you accept or overrule the first mover if they choose to invest 200 Ksh?

If overrule then how much would you invest?

The first mover and second mover will be determined at the end of the session if this task is selected for payment. You will also be asked to pick a ball at the end of the session to determine if the investment is doubled or lost.

3.C Men compared to women

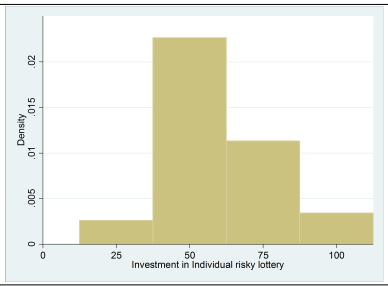
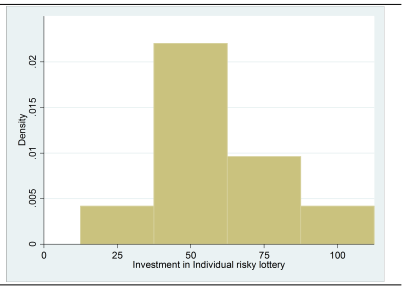
As mentioned in section 3.3.1, only married women and their spouses were eligible for these experiments. The decisions made by both women and men in the first task, the risk elicitation based on Gneezy and Potters (1997), are summarized in Figure 3.C.1. Contrary to several other studies (see e.g. Charness & Gneezy, 2012; Croson & Gneezy, 2009, among others), there are no statistically significant differences between the average investment decisions or the distribution of investment decisions made by men and women in this task. The t test is used to check for differences in means, and the Mann-Whitney test to check for differences in the distributions. [30] On average, both women and men invest more than half of their initial endowment in this individual lottery (approximately 60 Ksh and 59 Ksh, respectively).

However, there are differences when faced with the investment decision in the second task, where the endowment is doubled but the spouse now has a chance to change the investment. As shown in Figure 3.C.2, women invest less on average as the first mover compared to men (100 Ksh versus 108 Ksh by men), and this difference is statistically significant at the 1% level of significance based on a t test. [31]

[30] Croson and Gneezy (2009) highlight three possible explanations for the commonly observed gender differences in risk aversion: 1) women experience more intense nervousness and fear in response to negative outcomes leading them to be more risk averse in risky situations, 2) men are more overconfident in their success in uncertain situations than women, and 3) males are more likely to see risky situations as a challenge as opposed to women who see these as a threat. In this study setting the former two explanations may not hold and may explain the lack of observed gender difference in this task. In particular, livestock rearing, the main livelihood option in this region, is particularly risky due to drought, disease and rustlers. Such exposure to risk by women may dampen their nervousness and fear in response to negative outcomes and may also increase their confidence in their ability to overcome uncertain situations.

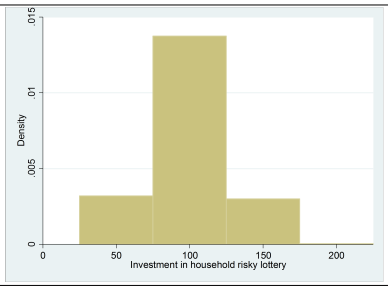
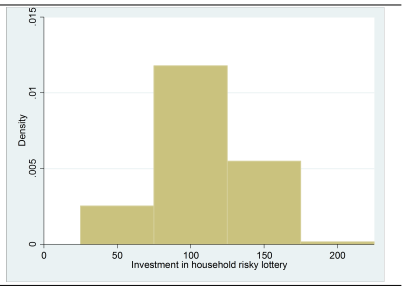
[31] The distribution of investments are also statistically different based on the Mann-Whitney test.

FIGURE 3.C.1: Summary of investments by women and men in individual risky lottery

	Women	Men
Mean	59.714	58.624
(Std Dev)	(18.392)	(20.125)
N	700	229
Distribution of investment decisions		
t test of differences in means (p-value)	0.468	
Mann-Whitney test (p-value)	0.339	

Note: Distribution graphs are on the same scale. The initial endowment given to participants to invest is 100 Ksh.

FIGURE 3.C.2: Summary of investments by women and men in household risky lottery

	Women	Men
Mean	99.786	108.297
(Std Dev)	(28.367)	(32.052)
N	700	229
Distribution of investment decisions		
t test of differences in means (p-value)	0.000	
Mann-Whitney test (p-value)	0.000	

Note: Distribution graphs are on the same scale.The initial endowment given to participants to invest is 200 Ksh. Two women and two men chose to invest the maximum of 200 Ksh.

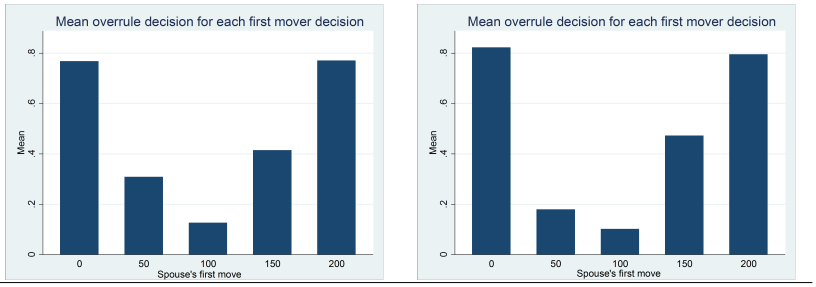
The decision of second movers to change the spouse's decisions (for all possible first mover decisions where $I_1^W \neq I_1^H$) are reported on in panel A of Figure 3.C.3. [32] On average, women change the spouse's decisions (0.542) more frequently than men (0.531), but this difference is small and not statistically significant (using both the *t* test and Mann-Whitney test). In panel B it is shown that

[32] Additionally for women observations are excluded where $I_1^H > I_1^W > I_2^W$, $I_1^W > I_1^H > I_2^W$, $I_2^W > I_1^H > I_1^W$, or $I_2^W > I_1^W > I_1^H$ and for men where $I_1^W > I_1^H > I_2^H$, $I_1^H > I_1^W > I_2^H$, $I_2^H > I_1^W > I_1^H$, or $I_2^H > I_1^H > I_1^W$.

the second movers change the spouse's decisions more frequently when the first mover investment decision is at an extreme (i.e. 0 Ksh or 200 Ksh) and are less likely to change their spouse's decision if it lies between these two extremes, a result which is expected as both men and women prefer investments that are away from the extremes (see Figures 3.C.1 and 3.C.2).

FIGURE 3.C.3: Summary of the decisions of second movers to change the spouse's decisions.

	Women	Men
Panel A: Decision to change or overrule spouse's first move		
Mean	0.542	0.531
(Std Dev)	(0.498)	(0.499)
N	2619	865
t test of differences in means (p-value)	0.609	
Mann-Whitney test (p-value)	0.964	
Panel B: Decision to change summarized by first mover decisions		



Note: Graphs are on the same scale. The decision to change or overrule is a binary variable equal to one if the woman decides to change her spouse's first move, and zero otherwise. The sample excludes those observations where $I_1^W = I_1^H$. Additionally for women observations are excluded where $I_1^H > I_1^W > I_2^W$, $I_1^W > I_1^H > I_2^W$, $I_2^W > I_1^H > I_1^W$, or $I_2^W > I_1^W > I_1^H$ and for men where $I_1^W > I_1^H > I_2^H$, $I_1^H > I_1^W > I_2^H$, $I_2^H > I_1^W > I_1^H$, or $I_2^H > I_1^H > I_1^W$.

Summary statistics for the *compromise* variable for both women and men are presented in Table 3.C.1. On average, men choose to compromise with their spouse more often than women (0.623 for men compared to 0.598 for women) but, the difference is small and not statistically significant at conventional levels (based on both the *t* test and Mann-Whitney test). The experimental measure of empowerment, *power*, is summarised in panel B where it is shown that women are more likely to change more than half of their husband's decision to their own compared to men (0.266 for women versus 0.240 for men) but this difference is not statistically significant at conventional levels.

During the experiments husbands were also asked the same household decision making questions posed to women during the endline survey (and used to derive *HDMI1* and *HDMI2*). These questions ask respondents to identify who in the household makes decisions about the purchase of food, household items, livestock, and paying school and medical fees. Figure 3.C.4 presents a comparison

TABLE 3.C.1: Second mover decisions: summary statistics.

	Women	Men
Panel A: average *compromise*		
Mean	0.598	0.623
(Std Dev)	(0.313)	(0.292)
N	700	229
t test of differences in means (p-value)	0.261	
Mann-Whitney test (p-value)	0.410	
Panel B: Binary indicator of *power*		
Mean	0.266	0.240
(Std Dev)	(0.442)	(0.428)
N	700	229
t test of differences in means (p-value)	0.437	

Note: Compromise is a binary indicator equal to zero if the woman decides to stick with her original investment decision which differs from her spouse's decision. The variable is equal to one if she changes her spouse's decision to some amount not equal to her original investment or if she chooses to accept her spouse's decision. In Panel B we take the average of the compromise variable across all first mover decisions. This average varies from zero (if the woman never compromises) to one (if she always compromises). The variable *power* is coded as one if in more than 50% of decisions as the second mover, the woman does not *compromise* at all with her husband.

of husband and wife responses to these questions with the sample restricted to women whose husbands also attended the experiments. Tests of differences in the overall distribution of these decision making variables between men and women are conducted and these differences are all found to be statistically significant (see Table 3.C.2). On average, husbands report having more say in household decisions than their wives report them having, with the exception of decisions on the purchase of small household items where the opposite is true. There is also much higher disagreement between spouses when it comes to reports on who has the final say about the purchase of food, compared to the final say on other household decisions.

These findings on gender-based differences in reporting on household decision making differ from several earlier studies. In many cases it is found that husbands attribute more dominance in the decision making to the wife than the wife does to herself (see e.g. Ghuman, Lee, & Smith, 2006, and references within). These differences between husband and wife reports may be due to a number of reasons including differing cognitive understanding of the questions, random errors in reporting, or social desirability bias with respondents choosing to conform to the social norms of household decision making in their responses (Ghuman et al., 2006).

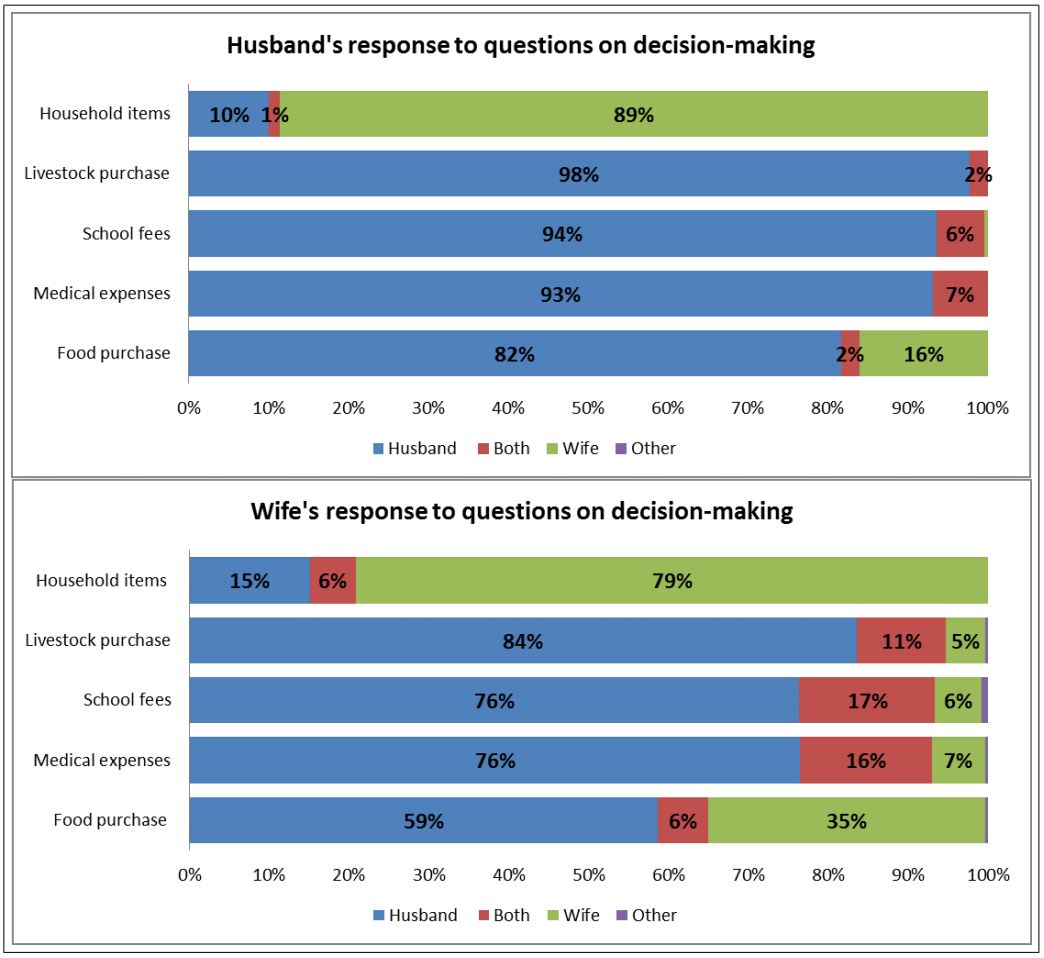


FIGURE 3.C.4: Comparison of husband and wife responses to survey questions on household decision making.

TABLE 3.C.2: Comparisons of self-reported household decision making by wives and their husbands

	(1) Wives	(2) Husbands	(3) % couples that disagree	(4) *t* test statistic	(5) Mann-Whitney test statistic
Purchasing small household items					
Mean	2.640	2.779			
(Std Dev)	(0.735)	(0.611)	28.4%	-2.177*	-2.502*
N	225	222			
Purchasing livestock					
Mean	1.209	1.032			
(Std Dev)	(0.510)	(0.194)	20.3%	4.821**	4.829**
N	225	222			
Paying for children's school fees					
Mean	1.286	1.072			
(Std Dev)	(0.569)	(0.283)	30.2%	5.065**	4.832**
N	224	222			
Paying for children's medical expenses					
Mean	1.298	1.068			
(Std Dev)	(0.585)	(0.253)	29.7%	5.405**	4.960**
N	225	222			
Purchasing food					
Mean	1.756	1.347			
(Std Dev)	(0.945)	(0.745)	45.5%	5.120**	5.047**
N	225	222			

Note: Each decision making variable is coded as follows: 1 = husband only, 2 = both husband and wife, 3 = wife only. The "other" category shown in Figure 3.C.4 are excluded from these estimates. Column (4) reports the t test statistic based on the following null hypothesis: mean (wife)- mean (husband) = 0. Column (5) reports the Mann-Whitney test statistic for the hypothesis that the two independent samples come from populations with the same distribution. *, ** denote significance at the 5% and 1% level of significance, respectively.

3.D Propensity score estimates

TABLE 3.D.1: Estimation of the propensity score

Variable	Coefficient estimate
1^{st} cycle of REAP	0.372** (0.168)
Literacy	-1.065** (0.537)
Business experience	0.493 (0.302)
Age	0.003 (0.016)
Age of spouse	-0.022 (0.014)
First wife in polygamous marriage	0.341 (0.215)
Second wife in polygamous marriage	0.239 (0.243)
HDMI1	-0.258** (0.108)
Household size	-0.297** (0.148)
# children in household (logged)	1.191* (0.691)
Proportion of school aged children in school	-3.358*** (0.971)
Proportion of school aged children in school (squared)	3.355*** (0.965)
Food expenditure per capita	-0.013** (0.006)
Non-food expenditure per capita	0.009 (0.006)
Total income per capita (logged)	-0.299*** (0.104)
Total savings per capita	0.008 (0.013)
Durable asset index	0.017 (0.035)
Durable asset index (squared)	-0.010 (0.006)
Durable asset index (cubed)	0.001 (0.000)
TLU per capita	-0.402 (0.661)
TLU per capita (squared)	0.517 (0.445)
TLU per capita (cubed)	-0.111 (0.080)
Benefited from HSNP cash transfer	-1.102** (0.448)
Participated in CARE savings groups	0.114 (0.341)
Location dummies	Yes
Pseudo R-squared	0.103
N	908
Likelihood Ratio chi-squared (36)	104.57

Note: Propensity score is estimated in Stata using –pscore– command with a logit model. The dependent variable equals one if the woman did not participate in the experiments, and zero otherwise. All monetary values are reported in 2014 USD, PPP terms. The *HDMI1* is standardised using the mean and the standard deviation of the overall sample of married women at baseline. Standard errors are shown in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels, respectively.

3.E Mean and distributional comparisons of empowerment measures

TABLE 3.E.1: Simple mean and distributional comparisons of empowerment measures between treatment and control groups

	power		*HDMI1*		*HDMI2*	
	Control	Treatment	Control	Treatment	Control	Treatment
Mean	0.247	0.286	0.015	-0.004	-0.001	-0.012
(Std Dev)	(0.432)	(0.452)	(1.014)	(0.978)	(0.998)	(1.007)
N	364	336	359	328	359	328
t test of differences in means						
(p-value)	0.251		0.802		0.885	
Mann-Whitney test						
(p-value)	0.250		0.950		0.767	

Note: The sample used for these comparisons are restricted to those women who participated in the experiments and is not adjusted to account for possible bias arising from self-selection into the experiment. The components of the *HDMI1* are coded as one if the woman has the final say in the decision or zero if otherwise whereas those of *HDMI2* are coded as one if the woman has the sole or joint final say in the decision or zero if otherwise. *HDMI1* and *HDMI2* are standardized using the control group mean and standard deviation. The variable *power* is coded as one if in more than 50% of decisions as the second mover, the woman does not compromise at all with her husband.

3.F REAP and the decision to *compromise*

TABLE 3.F.1: The impact of REAP on a woman's decision to *compromise* with her spouse.

	(1)	(2)	(3)	(4)
Treatment	-0.020	-0.028	0.009	-0.002
	(0.024)	(0.025)	(0.025)	(0.025)
Age		0.001		-0.000
		(0.002)		(0.002)
Age of spouse		-0.002		-0.002
		(0.002)		(0.002)
Literacy		0.017		-0.033
		(0.062)		(0.089)
Numeracy		0.043		0.076
		(0.046)		(0.048)
Wife in polygamous marriage		-0.000		0.041
		(0.027)		(0.027)
# children in household		-0.004		0.001
		(0.008)		(0.009)
# adults in household		0.038*		0.033
		(0.022)		(0.024)
TLU per capita		-0.024		-0.041*
		(0.022)		(0.022)
Durable asset index		-0.000		0.001
		(0.004)		(0.004)
Husband attends experiment		-0.026		-0.005
		(0.030)		(0.031)
Investment in individual risk task		0.000		0.000
		(0.001)		(0.001)
N	700	687	677	665
R-squared	0.166	0.183	0.163	0.185

Note: The dependent variable is the average of *compromise* across all first mover decisions of the spouse. Higher values of the dependent variable indicate less household bargaining power. Regressions include sub-location fixed effects. Demographic and household controls are at baseline levels. Regressions in columns 1 and 2 are unweighted. Regressions in columns 3 and 4 include the weights generated from the propensity score estimates. Standard errors are shown in parentheses. *, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

Chapter 4

Team heterogeneity and trust

4.1 Introduction

Teamwork is an integral part of the operation of many organisations, and teams have been increasingly used in interventions targeted at the poor. For example, over the last three decades we have witnessed the rise of the group lending model in microfinance, which in itself follows other experiences such as rotating savings and credit associations (ROSCAS). Teams can be more productive than individuals if members potentially bring a variety of skills, experience, knowledge and other factors to the production process. As such, diversity in teams is often argued to be one of the most influential determinants of teams' effectiveness (Page, 2007). But diversity may also lead to higher costs of communication and coordination, and result in more conflicts between individuals, which may be detrimental to performance (see, for example, Lazear (1999)).

A key consideration in understanding how team heterogeneity may affect teams' performance is through its effects on an individual's willingness to cooperate for the benefit the team when it might be individually optimal to free-ride on others' effort. Although much is known about the effect of institutions in overcoming the cooperation dilemmas in groups (Ostrom, 2010), Cárdenas, Chong, Ñopo, Horowitz, and Lederman (2009) suggest that more importance needs to be given to the behavioural aspects of the collective action problem, since in teams individuals may make decisions based on their sense of group affiliation, social distance or sympathy towards others in their team. In this paper we follow this line of enquiry, and examine the effect of team heterogeneity, measured as social distance (Akerlof, 1997), on trust and trustworthiness of individuals in a productive team.

Trust plays a central role in encouraging cooperation by reducing uncertainty and minimising the costs of limited information, and, within teams, increased trust is found to be positively related with team performance (see De Jong, Dirks, and Gillespie (2016) for a meta-analysis of studies relating trust to team performance). Trust allows teams to interact as if any uncertainty about team members were favourably resolved (De Jong & Elfring, 2010) which enables cooperation and

more effective and efficient collaboration (Dirks, 1999). A lack of trust, on the other hand, may instead lead individuals to focus on their own interests instead of that of the team (Joshi, Lazarova, & Liao, 2009). [1]

Numerous studies have found that both trust and trustworthiness are higher among individuals who are socially closer (see, for example, Glaeser, Laibson, Scheinkman, and Soutter (2000)). Greater heterogeneity in income, wealth, education or ethnicity may widen the social distance between individuals in a group, reducing the overall level of trust. A variety of mechanisms may explain this effect of heterogeneity on trust, but two have received particular attention in the *ex ante* literature: 1) diversity erodes the effectiveness of social control and sanctioning to enforce cooperative norms (Bernhard, Fehr, & Fischbacher, 2006); and, 2) diversity leads to the definition (or perception) of in-groups and out-groups, and in-group preferences leads to less trust in members of out-groups (Alesina & La Ferrara, 2002).

In this paper we take advantage of the exogenous formation of business groups to examine the effect of heterogeneity on trust. This exogenous assignment rules out the troublesome yet common self-selection issue in the team literature which arises due to the fact that in many team settings studied members self-select into teams Hansen, Owan, and Pan (2006). In November 2012 ultra-poor women in several locations across northern Kenya were identified as being eligible for the Rural Entrepreneur Access Project (REAP), a multifaceted approach to poverty alleviation. As part of REAP, eligible women were required to form firms composed of three women. All the inputs of the program (business skills training, a cash transfer to invest in a jointly run microenterprise, savings services, and on-going mentorship for a period of two years) were then directed at the group, rather than the individual beneficiaries. [2] [3] Although the original criteria for group formation by mentors specified that mentors were to place all eligible women into groups of three, in November 2012 eligible women in some locations were assigned to business groups by local staff whereas in others they were allowed to form their business groups from the set of eligible women in their location. [4] Given the limited time to form groups, the assignment of women to business groups was largely guided by the geographical proximity of women. As a result, in the groups formed exogenously it is unlikely that initial heterogeneity is correlated with other productive characteristics of the groups, including initial

[1] At the macro level, trust is considered essential for the creation and maintenance of economic prosperity (Fukuyama, 1995), regional development (Grootaert & van Bastelaer, 2002), collective action (Burt, 2001), and democratic governance (Putnam, 1993).

[2] Individuals, and not groups, decided whether or not to join savings associations formed as part of REAP.

[3] The program is implemented through a NGO, The BOMA Project. See `http://bomaproject.org/the-rural-entrepreneur-access-project/`

[4] The BOMA Project has since revised their selection criteria to allow members to form their groups from a list of eligible women.

levels of trust or trustworthiness. We take advantage of this exogenous assignment mechanism to explore the causal relationship between team heterogeneity and trust.

In June and July 2014 we played a modified version of the Berg, Dickhaut, and McCabe (1995) trust game with those REAP participants who received funding in the first and last funding cycles (March/April 2013 and March/April 2014, respectively). In the game, participants receive an endowment from which they can choose how much to send to one of their business partners, who would be randomly chosen and remains anonymous throughout the game. The amount sent is tripled and added to the partner's endowment who can then decide how much to return. The amount sent by the first mover is usually interpreted to measure trust while the amount returned to the trustor by the trustee is usually interpreted to measure trustworthiness.

We find that heterogeneity in groups does affect the level of trust and trustworthiness between business partners. In particular, differences in wealth (as measured by livestock ownership, the central asset in this environment) results in less trusting behaviour, with those who are wealthier than their expected partners less willing to transfer part of their endowment. This finding mimics social norms in these pastoral communities where poorer households tend to be excluded from informal livestock transfers due to their perceived inability to reciprocate (Illife, 1987). We also find that those who are poorer than their expected partners are less willing to transfer part of their endowment.

In the next section we discuss the study site in more detail, including the central role of livestock in defining social status and the formation of social networks. We also present more details about REAP, before proceeding to describe the trust game in Section 4.3.1. In Section 4.3.2 we turn to our measure of group heterogeneity which we define as a social distance that is estimated using a modification of the Euclidean norm proposed by Santos and Barrett (2010). In Section 4.4 we present our results before concluding in Section 4.5.

4.2 Study setting

This study was conducted in the southern and central parts of Marsabit County, in the Arid and Semi-Arid Lands (ASALs) of northern Kenya (see Figure 4.1), a region where more than 80% of the population are estimated to live below the national poverty line (Kenya National Bureau of Statistics and Society for International Development, 2013). The main livelihood option in these locations is pastoralism, with livestock serving both as a source of income and food for herders and their families. The transfer of livestock forms the material basis of

complex social networks, which can be mobilized in times of need (Little et al., 2008; McPeak, 2006).

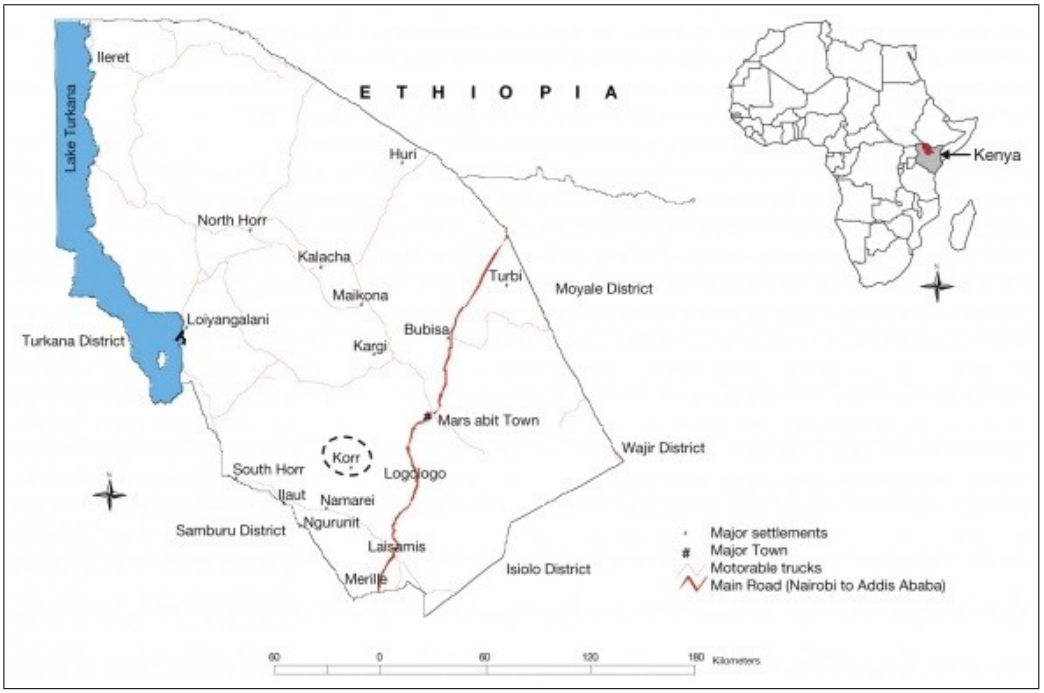


FIGURE 4.1: Map of Marsabit County (Warui & Kshatriya, 2009).

The anthropological literature highlights three main interpretations of livestock transfers in African pastoral societies which, as noted by McPeak (2006) reflect: 1) an *ex post* risk sharing mechanism where households facing unexpected herd loss are transferred livestock to help them cope with their economic emergencies; 2) an *ex ante* risk management mechanism where, as noted previously, livestock is used as a means to develop or reinforce social networks that can be drawn upon in times of need; and, 3) a mechanism promoting asset redistribution from the wealthier to the poorer. [5] Empirical evidence supports this second interpretation of livestock transfers where livestock is transferred or loaned to those who the lender expects to have the future capacity to reciprocate (McPeak, 2006; Santos & Barrett, 2011). As a result, poorer households are excluded from these informal credit arrangements leaving them vulnerable to shocks (Illife, 1987).

Pastoralism is highly susceptible to weather shocks, and droughts frequently have devastating impacts on household's livelihoods (Silvestri et al., 2012), resulting in many households no longer being able to meet their basic needs due to the loss of herds from which it is hard to recover (Barrett & Santos, 2014; Lybbert

[5] Huysentruyt, Barrett, and McPeak (2009) propose an alternative interpretation for inter-household livestock transfers among East African pastoralists where transfers are interpreted as a self-interested action to encourage recipient households to migrate with the donor households for both security and resource appropriation reasons. As a result, the poorest households are excluded from transfers as their herd sizes are too small to migrate.

et al., 2004). Without informal insurance networks to fall back on, these households are forced into begging, unskilled wage labor, different forms of petty trade, and become reliant on food aid to meet their dietary needs.

Many approaches have been used in northern Kenya to help such households deal with these economic shocks including livestock restocking, food aid and cash transfers (see for example Little et al. (2008)). One recent intervention, the Rural Entrepreneur Access Project (REAP), employs a multifaceted approach that is closely modelled after the poverty graduation approach pioneered by BRAC. [6] As part of REAP, ultra-poor women are provided with multiple interventions including business skills training, cash transfers to start an enterprise, savings services and regular follow-up visits by local mentors for a period of two years. [7] A recent evaluation of the impacts of this program concludes that participation in REAP results in significant improvements in income, savings, asset accumulation and food security, with the increase in income and savings being driven by women's participation in microenterprises and savings groups Gobin, Santos, and Toth (2016).

The program requires women to work in groups of three to run an enterprise. In this study we focus on a sub-sample of the ultra-poor women included in the Gobin, Santos, and Toth (2016) study. In particular, we look at those who received funding in March/April 2013 and March/April 2014. These women were initially identified as being eligible for REAP in November 2012 by a local selection committee. In some locations these women were assigned to groups by the local mentor whereas in other locations they were asked to form these groups with other eligible women. We focus our attention on those locations where women were assigned to groups by their mentors. In these locations group formation was based largely on geographical proximity, with women from the same village (or manyatta) often being placed into the same business group. The assignment of women to groups by the mentor likely leads to exogenous variation in the characteristics of members in a business group. This allows us to examine how within group diversity (reflecting the heterogeneity of members) affects group outcomes.

We focus our attention on the level of trust and trustworthiness within groups, which is seen to be key for group cooperation and is also likely to be affected by the heterogeneity of group members. Incentivised decision making tasks designed to measure trust and trustworthiness were implemented in June/July 2014

[6] The poverty graduation approach was pioneered by BRAC through the program Challenging the Frontiers of Poverty Reduction – Targeting the Ultra-Poor (CFPR/TUP) (Goldberg & Salomon, 2011; Matin et al., 2008). During a limited period (two years), participants of the CFPR/TUP program benefit from a set of interventions which includes initial consumption support and an asset transfer, together with savings services, skills training, and regular follow-up visits.

[7] Business mentors are specific to a location and are responsible for providing training and follow-up support to REAP beneficiaries. Each location comprises many sub-locations which are formed by smaller villages, known as manyattas.

in addition to other tasks designed to measure empowerment, risk and time preferences as well as group coordination (see Gobin, Santos, and Leibbrandt (2016) for details). In the next section we provide details of the trust experiment in which women who received funding in March/April 2013 or March/April 2014, and belonged to one of the nine locations where all business groups were assigned members by the mentor, participated.

4.3 Measuring trust, trustworthiness and heterogeneity

4.3.1 Trust and trustworthiness

We use a modified version of the original trust game (also known as the investment game) proposed by Berg et al. (1995) to derive a measure of trust and trustworthiness. In our version of the game, participants were anonymously and randomly paired with one of their business group members. [8] Both players received an endowment equivalent to 6000 experimental Ksh where 100 experimental Ksh was worth 1 Ksh. [9] The first mover could send any amount $k \in 0, 1000, 2000, 3000, 4000, 5000, 6000$ to the second mover. The amount sent was tripled by the experimenter after which the second mover could send back any amount between 0 and amount in hand (i.e, after the original transfer had been tripled). We used the strategy method to elicit each participant's trust and reciprocity decisions, and participants made decisions both in the role of the first mover and second mover. As the second mover, participants made conditional decisions for each of the following information sets: 1000, 2000, 3000, 4000, 5000, 6000 Ksh.

The trust game was part of a sequence of tasks (see Gobin, Santos, and Leibbrandt (2016) for details). If this task was chosen for payment then a random mechanism was used to determine which participant in the pair would be the first mover and the second mover. This ensured that the task was incentive compatible.

The trust game was also framed in an effort to make the experiment, and in particular the tripling of money by the experimenter, be more realistic and less abstract to participants. In the contextualised game participants could come from one of two communities; in one community goats are worth three times more than in the other community (3000 Ksh vs 1000 Ksh) due to the presence of a livestock market. Participants were then endowed with goats such that the total value of the

[8] Participants were aware that they were paired with a business partner but they were not aware of which of their two business partners formed the pair.

[9] At the time of the experiment the exchange rate was 1 USD to 85 Ksh and the average daily wage for menial labor was approximately 200 Ksh. The average daily consumption per capita for the study sample in March 2014 was approximately 55 Ksh.

endowment is the same in the two communities i.e. 6000 Ksh. [10] They were told that the goats were identical and would be sold at the next available opportunity. Participants in the community with a lower value for goats then made decisions about how many goats (out of six available) to transfer to their partner. The partner would then sell all of her animals and decide on how much of the proceeds to return to the first mover. The instructions used in this task are presented in Appendix 4.A.

The task was run separately in the various locations by a team of four enumerators and a research assistant who was in charge of overseeing all experimental procedures. [11] Invitations were sent to participants two weeks before the experiment and on the day before the experiment mentors reminded all eligible persons of the experiment. The experiments took place in churches, schools or meeting halls which were divided into separate spaces for each of the enumerators and the research assistant.

On the day of the experiment participants gathered outside of the location and the mentor was instructed to send participants into the venue. Once the women consented to taking part in the experiment enumerators began introducing the tasks. The enumerators stressed that payments would take place in private at the end of the sessions and that their decisions would not be revealed to other participants. The trust game was explained using props and a series of examples were used to further explain the game. This was followed by a series of questions to gauge the participant's understanding of the game. The explanations were repeated until it was clear that the game was understood by the participant. There was no communication between participants during the experiments.

Once an individual had completed this task (as well as any other tasks), she proceeded to determine her payoffs. [12] If the trust game was selected for payment then a coin toss was used to determine if she would be the first mover or second mover as well as which of her business partners' decisions would be used to determine her payoffs. If her business partners had not yet completed the tasks and the trust game was selected for payment, then she was asked to wait until they had finished. [13] If neither of her business partners showed up during the course of the day then she was not rewarded for this task.[14]

[10] In the role of the sender, participants receive 6 goats valued at 1000 Ksh each and in the role of the sender they receive 2 goats valued at 3000 Ksh each.

[11] Eight local language enumerators underwent a three day training session on the experimental procedures before being divided into two teams.

[12] The tasks selected for payoffs were determined randomly by choosing numbered balls from a bag.

[13] Participants who had completed the tasks were kept separate from those who were still awaiting their turn.

[14] Prior to beginning the task the participant was asked if her business partners would be attending, if she reported that they would attend or that she was uncertain then she was allowed to participate in the trust game.

The measure of trust is derived by dividing the value of the goats sent (in the first mover's location) by the value of the endowment. The measure of reciprocity or trustworthiness is the amount returned by the second mover as a proportion of the amount available to return. When measured this way, both trust and trustworthiness are proportions, falling between zero and one. The Nash Equilibrium if participants have self-regarding preferences is for the second mover to keep the entire amount and in anticipation of this behaviour the first mover transfers nothing. On the other hand the socially optimum strategy is for the first mover to transfer the entire amount (as it earns a 300% interest). Replications of the trust game around the world consistently find that individuals are willing to send and receive positive amounts although there is considerable variation in the both the amount sent and the amount returned (see Johnson and Mislin (2011)).

Of the 690 REAP beneficiaries who received funding in March/April 2013 or March/April 2014 and were assigned to business groups by their mentors, 507 participated in the trust experiments. [15] Figure 4.2 presents the decisions made by participants as both the first mover and second mover in the trust experiments. On average, participants reveal a willingness to trust, with first movers sending 44% of their endowment and, although there is variation in the amount sent, less than 1% choose to send zero and less than 2% send the entire endowment (as shown in the histogram in panel A of Figure 4.2). We also find that, on average, participants are trustworthy and return higher amounts if the first mover sends more. [16]

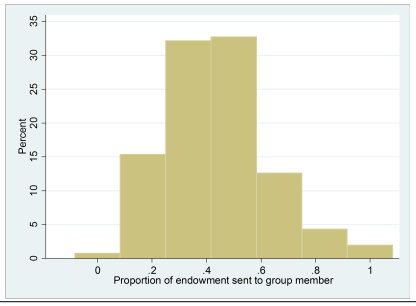
The use of the strategy method results in six observations of trustworthiness for each participant i.e. one decision for each (hypothetical) amount sent by the first mover. To obtain a single measure of trustworthiness we follow the approach of Altmann, Dohmen, and Wibral (2008) and estimate, for each participant, an OLS regression of the amounts returned by the second mover on the (hypothetical) amounts sent by the first mover, forcing the slope through zero. The resulting slope coefficient is then used as our measure of trustworthiness with higher values indicating higher levels of trustworthiness. If the second mover always matches her final payoff with that of her business partner then her trustworthiness coefficient is two. We summarise this measure of trustworthiness in panel C of Figure 4.2 which shows that a large proportion of participants (more than 49%) have a trustworthiness coefficient greater than two. Additionally, more than 93% of participants have a slope coefficient greater than one, which means that these women leave their business partners with a positive return on their trust i.e. they are returned more than the amount they sent.

[15] Of these 507 women, 233 received funding from REAP in March/April 2013 and 274 in March/April 2014.

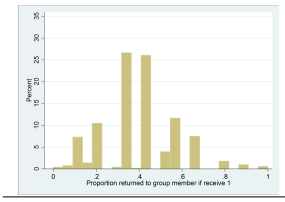
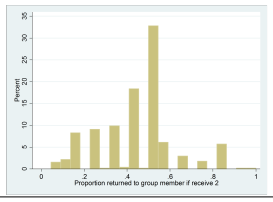
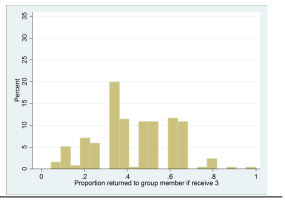
[16] The behaviour of participants in this trust game (in terms of the proportion sent and returned) is largely in line with the behaviour of participants in other studies that use the trust game with non-student samples (see, for example, Johnson and Mislin (2011) for a useful meta-analysis of these studies).

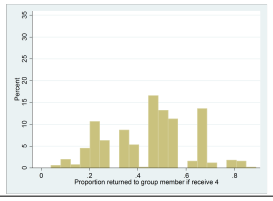
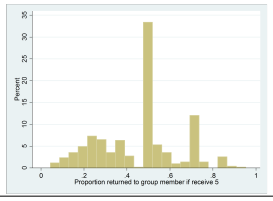
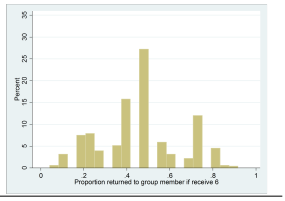
FIGURE 4.2: Summary of trust and trustworthiness

Panel A: Trust - proportion of endowment sent

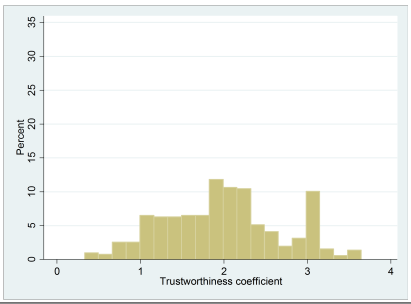
Mean	0.437
(Std Dev)	(0.193)
Distribution	

Panel B: Trustworthiness - proportion returned

	Receive 1 goat	Receive 2 goats	Receive 3 goats
Mean	0.403	0.438	0.431
(Std Dev)	(0.170)	(0.175)	(0.175)
Distribution			

	Receive 4 goats	Receive 5 goats	Receive 6 goats
Mean	0.443	0.450	0.471
(Std Dev)	(0.174)	(0.186)	(0.193)
Distribution			

Panel C: Trustworthiness coefficient

Mean	1.984
(Std Dev)	(0.690)
Distribution	

Note: The y-axis of the various histograms are scaled the same. The sample size is 507 women except for the proportion returned when the second mover receives 2 or 5 goats where the sample size is 506. The trustworthiness coefficient is based on a regression of the amount returned on the amount sent

4.3.2 Heterogeneity in groups

In the trust game participants are paired with one of their business partners who is chosen randomly and remains anonymous throughout the task. Participants are made aware that a coin toss determines which business partner is chosen and

are equally likely to be paired with either of their business partners. Participants would therefore consider the expected characteristics of their business partners relative to themselves in making their decisions both as the first mover and the second mover in the trust game. This difference between participants and their business partners can be expressed in terms of social distance.

To capture the social distance between participants and their business partners we follow the approach of Santos and Barrett (2010) who use a modification of the Euclidean norm to account for asymmetries in the effect of differences in observable characteristics of two parties. [17] For any continuous variable x, the distance between the participant, i, and her (expected) business partner, j, is given by the following two variables:

$$I(x_i - E(x_j) < 0) \times |x_i - E(x_j)| + I(x_i - E(x_j) \geq 0) \times |x_i - E(x_j)| \tag{4.1}$$

where $I_{(\bullet)}$ is an indicator function taking value one if true, zero otherwise. Under their approach, the distance between categorical variables is defined by a set of dummy variables that consider the several possible characterisations of the match. [18] This approach allows us to capture the directional effect of differences in the characteristics of participants and their business partners.

The characteristics used to measure team heterogeneity in our analysis are summarised in Table 4.1 and include characteristics such as asset ownership (including livestock and durable assets), which we interpret as indicators of wealth and, in the case of livestock, social status. We also include characteristics that capture an individual's skills and experience including her age, literacy, numeracy and business experience. Finally, in the absence of information on social networks and prior relations, we proxy for the closeness of individuals by considering their clan membership.

4.4 The effect of heterogeneity on trust and trustworthiness

The relationship between trust and trustworthiness and the heterogeneity of business groups is estimated using a linear regression model that includes the variables measuring differences presented in Table 4.1 as well as location fixed effects and a binary indicator that is equal to one if either of the woman's other business

[17] This is an important consideration given that for the participant it is likely that not only the absolute difference between herself and her business partners matters, but also her position relative to her business partners.

[18] In this paper many categorical variables are treated as continuous given that we are measuring the distance between the participant and the expected value of her business partners.

TABLE 4.1: Variable definitions and descriptive statistics

Variable	Definition	Mean (Std Dev)
Respondent is older	Absolute value of the age difference (in years) between respondent and the average age of her business partners, if she is older than this average, 0 otherwise.	4.584 (7.604)
Respondent is younger	Absolute value of the age difference (in years) between respondent and the average age of her business partners, if she is younger than this average, 0 otherwise.	4.461 (6.575)
Respondent is more literate	Absolute value of the difference in literacy between respondent and the average literacy of her business partners, if she is more literate than this average, 0 otherwise.	0.026 (0.155)
Respondent is less literate	Absolute value of the difference in literacy between respondent and the average literacy of her business partners, if she is less literate than this average, 0 otherwise.	0.021 (0.114)
Respondent is more numerate	Absolute value of the difference in numeracy between respondent and the average numeracy of her business partners, if she is more numerate than this average, 0 otherwise.	0.063 (0.186)
Respondent is less numerate	Absolute value of the difference in numeracy between respondent and the average numeracy of her business partners, if she is less numerate than this average, 0 otherwise.	0.056 (0.218)
Respondent has more business experience	Absolute value of the difference in business experience between respondent and the average business experience of her business partners, if she is has more experience than this average, 0 otherwise.	0.052 (0.198)
Respondent has less business experience	Absolute value of the difference in business experience between respondent and the average business experience of her business partners, if she is has less experience than this average, 0 otherwise.	0.057 (0.196)
Respondent has more livestock	Absolute value of the difference in livestock (in TLU) between respondent and the average livestock of her business partners, if she has more than this average, 0 otherwise.	0.989 (2.034)
Respondent has less livestock	Absolute value of the difference in livestock (in TLU) between respondent and the average livestock of her business partners, if she has less than this average, 0 otherwise.	0.975 (1.823)
Respondent has more durable assets	Absolute value of the difference in durable assets between respondent and the average durable assets of her business partners, if she has more than this average, 0 otherwise.	1.226 (2.323)
Respondent has less durable assets	Absolute value of the difference in durable assets between respondent and the average durable assets of her business partners, if she has less than this average, 0 otherwise.	1.138 (1.930)
All members from different clans	Dummy variable, equal to 1 if respondent and business partners belong to different clans.	0.168 (0.374)
Member is from different clan but other two are from same	Dummy variable, equal to 1 if respondent is from a different clan but business partners belong to the same clan.	0.087 (0.282)
Member is from same clan as one other member	Dummy variable, equal to 1 if respondent is from the same clan as only one other business partner.	0.176 (0.381)

Note: The sample is restricted to those 507 women who participated in the trust experiment and were assigned to business groups by their mentor. All variables are measured at baseline. Literacy, numeracy and business experience are dummy variables equal to one if the individual if literate, numerate or has business experience. Livestock is measured in Tropical Livestock Units (TLU) where one TLU is equivalent to 1 head of cattle, 0.7 camels, 10 sheep/goats, or 2 donkeys. Durable asset ownership is measured using an index which is described in Gobin, Santos, and Toth (2016).

partners attend the experiment. The estimates of these regressions with the proportion sent (trust) and the trustworthiness coefficient as the dependent variables are reported in Table 4.2. In this table we also report on two other specifications (for each dependent variable) that include the funding cycle of the participant as

a control variable. This allows us to control for any learning effects that might affect the level of trust and trustworthiness between groups that had worked together for 15 months (March/April 2013 funding cycle) compared to those that worked together for 3 months (March/April 2014 funding cycle). In the analysis that follows we discuss our preferred specification which includes both our indicator variable for funding cycle as well as variables capturing the social distance between business group members (columns (3) and (6)). [19]

We first examine the relationship between social distance and our measure of trust. We find that most variables expressing differences in members are not statistically significant at conventional levels with the notable exception of livestock. The lack of evidence of a significant association between many of these measures and trust may be due to group members being similar to each other, with any differences between them not reflecting a large enough social distance to have an effect on trust. Alternatively, these business groups have been interacting with each other repeatedly for at least three months prior to the trust experiment. [20] These interactions may allow for learning about group members and, for example, reduce the effect of differences in characteristics such as clan membership or age on trust.

As noted previously, we find a statistically significant effect of differences in livestock ownership on trust. Participants who own either more or less livestock than their (average) business partner are found to be less trusting of them. This effect of differences in livestock ownership on trust likely reflects the importance of livestock as a signal of social status in our study sites. [21]

The anthropological literature, as well as other empirical studies, highlight the exclusion of livestock-poor households from the social networks and informal credit arrangements (through the transfer of livestock) due to their perceived inability to reciprocate in the future (McPeak, 2006; Santos & Barrett, 2011). Our trust game which is framed in terms of livestock transfers, may lead participants to think about their business partners in terms of their livestock holdings which may make this characteristic more salient than others, reinforcing the importance of differences in terms of livestock ownership over the importance of other variables. Participants with more livestock, perhaps in anticipation of lower levels of reciprocation from their livestock-poor partners, therefore transfer a smaller proportion of their endowment to them. On the other hand, participants with less

[19] In Appendix 4.B results are presented for specifications that includes two binary indicators for the number of business members that attend the experiment instead of one indicator for whether or not business partners attended. The findings on the effect of heterogeneity on trust and trustworthiness are unchanged.

[20] Recall that our sample began benefitting from REAP in March/April 2013 and the experiments were conducted in June/July 2014.

[21] We cannot reject the null hypothesis that the difference between the two coefficients is equal to zero.

TABLE 4.2: The effect of heterogeneity on trust and trustworthiness

	Trust			Trustworthiness		
	(1)	(2)	(3)	(4)	(5)	(6)
Mar/Apr 2013 funding cycle	0.002		0.005	-0.005		0.002
	(0.009)		(0.009)	(0.029)		(0.030)
Business partners attend	0.032	0.032	0.030	-0.052	-0.059	-0.058
	(0.035)	(0.035)	(0.036)	(0.104)	(0.111)	(0.112)
Respondent is older		0.001	0.001		-0.008**	-0.008**
		(0.001)	(0.001)		(0.004)	(0.004)
Respondent is younger		-0.001	-0.001		-0.004	-0.004
		(0.001)	(0.001)		(0.005)	(0.005)
Respondent is more literate		0.012	0.008		-0.235	-0.237
		(0.056)	(0.056)		(0.236)	(0.240)
Respondent is less literate		0.061	0.058		-0.195	-0.196
		(0.051)	(0.050)		(0.242)	(0.241)
Respondent is more numerate		-0.044	-0.044		0.074	0.074
		(0.052)	(0.052)		(0.175)	(0.175)
Respondent is less numerate		-0.032	-0.032		0.109	0.109
		(0.045)	(0.045)		(0.153)	(0.153)
Respondent has more business experience		0.054	0.056		-0.006	-0.005
		(0.052)	(0.052)		(0.136)	(0.136)
Respondent has less business experience		-0.014	-0.012		0.093	0.094
		(0.043)	(0.042)		(0.153)	(0.153)
Respondent has more livestock		-0.010**	-0.010**		-0.016	-0.016
		(0.004)	(0.004)		(0.012)	(0.012)
Respondent has less livestock		-0.011**	-0.011**		0.002	0.002
		(0.004)	(0.004)		(0.018)	(0.018)
Respondent has more durable assets		-0.004	-0.005		-0.006	-0.006
		(0.003)	(0.003)		(0.015)	(0.015)
Respondent has less durable assets		-0.005	-0.005		-0.004	-0.004
		(0.004)	(0.004)		(0.018)	(0.018)
All members from different clans		0.051*	0.053*		-0.088	-0.087
		(0.031)	(0.031)		(0.107)	(0.107)
Member is from different clan but other two are from same		-0.032	-0.031		0.061	0.062
		(0.029)	(0.029)		(0.129)	(0.130)
Member is from same clan as one other member		-0.020	-0.019		0.115	0.115
		(0.025)	(0.025)		(0.081)	(0.081)
R-squared	0.027	0.069	0.069	0.132	0.152	0.152
N	507	507	507	507	507	507

Note: All regressions include location fixed effects. Robust standard errors clustered at the business group level are reported in parentheses. Trust is measured by the proportion of the endowment sent by first movers to their business partners. The measure of trustworthiness is based on the coefficients obtained from a regression of the amounts returned by the second mover on the (hypothetical) amounts sent by the first mover. See Table 4.1 for a description of explanatory variables. For the three clan dummy variables, the reference category is all members are from the same clan. *, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

livestock, in anticipation of receiving smaller reciprocal transfers from their relatively better off business partners, send a smaller proportion of their endowment as the first mover.

We do not ask participants how much they expect their business partners to return so we are unable to test if our hypotheses about their expectations are correct. However, we do observe the second mover behaviour of all participants, i.e. their trustworthiness. The results of regressions with our measure of trustworthiness as the dependent variable are presented in columns (4) to (6) of Table 4.2. Again,

our preferred specification is the one that includes all social distance variables as well as the indicator for assignment to funding cycle (column (6)). These results show that participants who own more livestock than their business partners are less trustworthy on average whereas those who own less livestock than their partners are more trustworthy. However, neither effect is statistically significant at conventional levels. Similarly, we see that most measures of social distance are not significantly associated with our measure of trustworthiness, and here again it is possible that this reflects a lack of variation in the measures of social distance. However, we do find that participants who are older than their business partners are less trustworthy on average (significant at the 5% level), a finding that goes against that of other studies which show that older adults are no different in their trustworthiness compared to young adults when interacting with other-age partners in a trust game (see, for example, Bailey et al. (2015)). [22]

4.5 Conclusion

In this study we estimated the effect of team heterogeneity on trust and trustworthiness and showed that many measures of social distance are not associated with the decisions made in the trust game, which is likely due to the lack of variation in many of these measures in our sample. However, we do find that differences in livestock ownership has a negative and significant effect on the amount sent in a trust game but not on the amount returned. This leads us to conclude that the decisions in the game are guided by societal norms where the poor are excluded from many important social networks and are less likely to benefit from informal livestock transfers due to a perceived inability to reciprocate, a mechanism that is corroborated by a significant amount of prior literature in economic anthropology.

This paper is the first step in the analysis of heterogeneity on team performance. In future work we will extend the analysis to look at overall indicators of performance (such as returns to the business), in addition to alternative mechanisms along the causal chain (such as free-riding and conflicts). We will also address selection issues in the sample of women who attended the experiments so that we are able to identify the effects of learning on the decisions made in the trust game.

[22] Bailey et al (2015) is the only other study that we have found that looks at the effect of differences in age between participants on their decisions in a trust game.

Appendix

4.A Instructions for trust game

In this part of the experiment you will be required to make one decision as an individual. You will be randomly paired with one of your business partners and each one of you will be randomly assigned to a specific type in the group, designated as Person A, and Person B. All persons will remain anonymous during this task.

Suppose Person A and Person B live in different locations.

Person A receives 6 goats from an asset transfer program and in her location goats are worth 1000 KSH.

Person B receives 2 goats from the same program but in her location goats are worth 3000 KSH.

The goats are identical but are worth more in Person B's location due to the presence of a livestock market.

Person A and Person B both prefer to have cash and will sell the goats in their respective locations at the next opportunity.

Before selling their goats, Person A can choose to send any amount of goats between 0 and 6 to Person B.

The goats sent by Person A triple in value when they get to Person B who can sell them each for 3000 KSH instead of 1000 KSH.

Person B will sell all the goats she receives from the transfer program as well as any goats she receives from Person A.

Person B can then decide to send any amount of money between 0 KSH and the amount she sells her 2 goats for plus the amount she sells the goats that were sent to her by Person A.

EXAMPLE

Suppose Person A chooses to send 3 goats to Person B. Person B sells her 2 goats plus the 3 goats sent by Person A and gets 15000 KSH.

Person B then decides to keep 13000 KSH and send 2000 KSH to Person A.

Person A kept 3 goats and ends up with 3000 KSH from the sale of those 3 goats that she did not send to Person B plus the 2000 KSH sent by Person B (for a total of 5000 KSH).

Person B ends up with 6000 KSH from her initial 2 goats from the transfer program and 7000 KSH from the extra money she received from selling the 3 goats sent by Person A (for a total of 13000 KSH).

NOTE, For this task you will be rewarded 1% of the amount that was actually transacted. So Person A will receive 50 Ksh and Person B will receive 130 Ksh.

Question:

You are Person A and you decide to send 5 goats to Person B ...

What is the maximum amount Person B can return to you?

Now suppose Person B returns 3000 KSH to you ...

How much will you receive in total?

How much will Person B receive in total?

If you are PERSON A ...

How many goats will you send to PERSON B?

If you are PERSON B ...

Suppose PERSON A has sent 1 goat to you. You now have 3 goats which you sell for 3000 KSH each and you end up with a total of 9000 KSH. How much money between 0 KSH and 9000 KSH will you send to Person A?

Suppose PERSON A has sent 2 goats to you. You now have 4 goats which you sell for 3000 KSH each and you end up with a total of 12000 KSH. How much money between 0 KSH and 12000 KSH will you send to Person A?

Suppose PERSON A has sent 3 goats to you. You now have 5 goats which you sell for 3000 KSH each and you end up with a total of 15000 KSH. How much money between 0 KSH and 15000 KSH will you send to Person A?

Suppose PERSON A has sent 4 goats to you. You now have 6 goats which you sell for 3000 KSH each and you end up with a total of 18000 KSH. How much money between 0 KSH and 18000 KSH will you send to Person A?

Suppose PERSON A has sent 5 goats to you. You now have 7 goats which you sell for 3000 KSH each and you end up with a total of 21000 KSH. How much money between 0 KSH and 21000 KSH will you send to Person A?

Suppose PERSON A has sent 6 goats to you. You now have 8 goats which you sell for 3000 KSH each and you end up with a total of 24000 KSH. How much money between 0 KSH and 24000 KSH will you send to Person A?

4.B Alternative regression specifications

TABLE 4.B.1: The effect of heterogeneity on trust and trustworthiness

	Trust			Trustworthiness		
	(1)	(2)	(3)	(4)	(5)	(6)
Mar/Apr 2013 funding cycle	0.001		0.006	0.000		0.014
	(0.019)		(0.019)	(0.060)		(0.061)
2 business members attend	0.042	0.041	0.041	-0.081	-0.086	-0.085
	(0.037)	(0.037)	(0.037)	(0.111)	(0.118)	(0.118)
3 business members attend	0.025	0.024	0.026	-0.029	-0.041	-0.037
	(0.037)	(0.036)	(0.036)	(0.108)	(0.114)	(0.115)
Respondent is older		0.001	0.001		-0.008**	-0.008**
		(0.001)	(0.001)		(0.004)	(0.004)
Respondent is younger		-0.001	-0.001		-0.004	-0.004
		(0.001)	(0.001)		(0.005)	(0.005)
Respondent is more literate		0.011	0.009		-0.233	-0.239
		(0.056)	(0.057)		(0.233)	(0.238)
Respondent is less literate		0.059	0.058		-0.191	-0.195
		(0.052)	(0.051)		(0.244)	(0.243)
Respondent is more numerate		-0.040	-0.040		0.063	0.062
		(0.052)	(0.052)		(0.175)	(0.175)
Respondent is less numerate		-0.028	-0.029		0.098	0.098
		(0.045)	(0.045)		(0.152)	(0.152)
Respondent has more business experience		0.055	0.056		-0.008	-0.001
		(0.052)	(0.052)		(0.137)	(0.138)
Respondent has less business experience		-0.015	-0.014		0.095	0.098
		(0.043)	(0.043)		(0.153)	(0.153)
Respondent has more livestock		-0.010**	-0.010**		-0.016	-0.017
		(0.004)	(0.004)		(0.012)	(0.012)
Respondent has less livestock		-0.011**	-0.011**		0.002	0.001
		(0.004)	(0.004)		(0.018)	(0.018)
Respondent has more durable assets		-0.005	-0.005		-0.005	-0.005
		(0.003)	(0.003)		(0.015)	(0.015)
Respondent has less durable assets		-0.005	-0.005		-0.003	-0.003
		(0.004)	(0.005)		(0.018)	(0.018)
All members from different clans		0.050	0.052*		-0.086	-0.083
		(0.030)	(0.030)		(0.107)	(0.107)
Member is from different clan but other two are from same		-0.033	-0.032		0.065	0.066
		(0.029)	(0.029)		(0.128)	(0.128)
Member is from same clan as one other member		-0.022	-0.021		0.118	0.120
		(0.025)	(0.025)		(0.081)	(0.081)
R-squared	0.028	0.070	0.071	0.133	0.153	0.153
N	507	507	507	507	507	507

Note: All regressions include location fixed effects. Robust standard errors clustered at the business group level are reported in parentheses. Trust is measured by the proportion of the endowment sent by first movers to their business partners. The measure of trustworthiness is based on the coefficients obtained from a regression of the amounts returned by the second mover on the (hypothetical) amounts sent by the first mover. See Table 4.1 for a description of explanatory variables. For the two business member attendance dummy variables the reference category is only one member attends. For the three clan dummy variables, the reference category is all members are from the same clan. *, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

Bibliography

Akerlof, G. (1997). Social distance and social decisions. *Econometrica*, *65*(5), 1005–1027.

Alesina, A., & La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, *85*(2), 207–234.

Allen, H. (2006). Village Savings and Loans Associations: sustainable and cost-effective rural finance. *Small Enterprise Development*, *17*(1), 61–68.

Almas, I., Armand, A., Attanasio, O., & Carneiro, P. (2015). *Measuring and changing control: Women's empowerment and targeted transfers* (Working Paper No. 21717). Cambridge, MA: National Bureau of Economic Research.

Altmann, S., Dohmen, T., & Wibral, M. (2008). Do the reciprocal trust less? *Economics Letters*, *99*(3), 454–457.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495.

Bailey, P., Slessor, G., Rieger, M., Rendell, P. G., Moustafa, A., & Ruffman, T. (2015). Trust and trustworthiness in young and older adults. *Psychology and Aging*, *30*(4), 977–986.

Bandiera, O., Burgess, R., Das, N. C., Gulesci, S., Rasul, I., & Sulaiman, M. (2016). *Labor markets and poverty in village economies* (Economic Organisation and Public Policy Discussion Papers Series). London, UK: LSE, STICERD.

Banerjee, A. V., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parient, W., ... Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, *348*(6236), 1260799.

Banerjee, A. V., Karlan, D., & Zinman, J. (2015). Six randomized evaluations of microcredit: introduction and further steps. *American Economic Journal: Applied Economics*, *7*(1), 1–21.

Barrett, C. B., & Santos, P. (2014). The impact of changing rainfall variability on resource-dependent wealth dynamics. *Ecological Economics*, *105*, 48–54.

Bauchet, J., Morduch, J., & Ravi, S. (2015). Failure vs displacement: Why an innovative anti-poverty program showed no net impact in South India. *Journal of Development Economics*, *116*, 1–16.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*, 289–300.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122-142.

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *American Economic Review*, *96*(2), 217–221.

Bianchi, M., & Bobba, M. (2013). Liquidity, risk, and occupational choices. *Review of Economic Studies*, *80*(2), 491–511.

Blattman, C., Green, E., Jamison, J., Lehmann, C., & Annan, J. (2016). The returns to microenterprise development among the ultra-poor: A field experiment in post-war Uganda. *American Economic Journal: Applied Economics*, *8*(2), 35–64.

Booysen, F., van der Berg, S., Burger, R., von Maltitz, M., & du Rand, G. (2008). Using an asset index to assess trends in poverty in seven Sub-Saharan African countries. *World Development*, *36*(6), 1113–1130.

Braaten, R., & Martinsson, P. (2015). *Experimental measures of household decision power* (Working Paper No. 02/2015). Oslo, Norway: CREE.

BRAC. (2013). *An end in sight for extreme poverty. scaling up BRAC's graduation model for the ultra-poor* (Briefing Note # 1: Ending Extreme Poverty). New York, NY: BRAC USA. Retrieved from http://tup.brac.net/images/BRAC_Briefing_Document_on_TUP.pdf

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, *14*(3), 375–398.

Brosig, J., Weimann, J., & Yang, C. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics*, *6*(1), 75–90.

Bruhn, M. (2011). License to sell: the effect of business registration reform on entrepreneurial activity in Mexico. *The Review of Economics and Statistics*, *93*(1), 382–386.

Bruhn, M., Karlan, D., & Schoar, A. (2013). *The impact of consulting services on small and medium enterprises: evidence from a randomized trial in Mexico* (Policy Research Working Paper Series No. 6508). Washington, DC: World Bank.

Burt, R. (2001). Structural holes versus network closure as social capital. In N. Lin, K. Cook, & R. Burt (Eds.), *Social capital: Theory and research.* New York: Aldine de Gruyter.

Cárdenas, J., Chong, A., Ñopo, H., Horowitz, A., & Lederman, D. (2009). To what extent do Latin Americans trust, reciprocate, and cooperate? Evidence from experiments in six Latin American countries. *Economica*, *9*(2), 45–94.

Carlsson, F., Martinsson, P., Qin, P., & Sutter, M. (2013). The influence of spouses on household decision making under risk: An experiment in rural China. *Experimental Economics*, *16*(3), 383–401.

Carter, J., Byrne, S., Schrader, K., Kabir, H., Uraguchi, Z., Pandit, B., ... Fendrich, P. (2014). Learning about women's empowerment in the context of development projects: Do the figures tell us enough? *Gender and Development*,

22(2), 327–349.

Carter, M. R., Tjernström, E., & Toledo, P. (2016). *Heterogeneous impact dynamics of a rural business development program in Nicaragua* (Working Paper No. 22628). Cambridge, MA: National Bureau of Economic Research.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior and Organization*, *83*(1), 50–58.

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior and Organization*, *87*, 43–51.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–474.

Dahl, R. (1957). The concept of power. *Behavioural Science*, *2*(3), 201–215.

De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*, *101*(8), 1134–1150.

De Jong, B. A., & Elfring, T. (2010). How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring and effort. *Academy of Management Journal*, *53*(3), 535–549.

de Mel, S., McKenzie, D., & Woodruff, C. (2008). Returns to capital in microenterprises: evidence from a field experiment. *Quarterly Journal of Economics*, *123*(4), 1329–1372.

de Palma, A., Picard, N., & Ziegelmeyer, A. (2011). Individual and couple decision behavior under risk: Evidence on the dynamics of power balance. *Theory and Decision*, *70*(1), 45–64.

Dirks, K. T. (1999). The effects of interpersonal trust on work group performance. *Journal of Applied Psychology*, *84*(3), 445–455.

Drexler, A., Fischer, G., & Schoar, A. (2014). Keeping it simple: financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, *6*(2), 1–31.

Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic Literature*, *50*(4), 1051–1079.

Duvendack, M., Palmer-Jones, R., & Vaessen, J. (2014). Meta-analysis of the impact of microcredit on women's control over household decisions: Methodological issues and substantive findings. *Journal of Development Effectiveness*, *6*(2), 73–96.

Eckel, C., & Grossman, P. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, *23*(4), 281–295.

Elliot, H., & Fowler, B. (2012). *Markets and poverty in northern Kenya: Towards a financial graduation model* (Tech. Rep.). Nairobi, Kenya: Financial Sector Deepening.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data or tears: An application to educational enrollments in states of

India. *Demography*, *38*(1), 115–132.

Fratkin, E. M. (2004). *Ariaal pastoralists of Kenya: studying pastoralism, drought, and development in Africa's arid lands* (Second ed.). Pearson Education, Incorporated.

Fukuyama, F. (1995). *Trust: The social values and the creation of prosperity*. New York: Free Press.

Ghuman, S., Lee, K., & Smith, H. (2006). Measurement of women's autonomy according to women and their husbands: Results from five Asian countries. *Social Science Research*, *35*(1), 1–28.

Gindling, T. H., & Newhouse, D. (2014). Self-employment in the developing world. *World Development*, *56*, 313–331.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, *115*(3), 811–846.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, *112*(2), 631–645.

Gobin, V., Santos, P., & Leibbrandt, A. (2016). *Not compromising: the impact of a poverty graduation program on women's empowerment* (Unpublished manuscript). Australia: Department of Economics, Monash Business School, Monash University.

Gobin, V., Santos, P., & Toth, R. (2016). *Poverty graduation with cash transfers: A randomised evaluation* (Discussion Paper No. 23-16). , Australia: Department of Economics, Monash Business School, Monash University.

Goldberg, N., & Salomon, A. (2011). *Ultra Poor Graduation Pilots: Spanning the gap between charity and microfinance* (Commissioned Workshop Paper). Valladolid, Spain: 2011 Global Microcredit Summit.

Grootaert, C., & van Bastelaer, T. (2002). *The role of social capital in development: An empirical assessment*. Cambridge, UK: Cambridge University Press.

Hansen, Z., Owan, H., & Pan, J. (2006). *The impact of group diversity on performance and knowledge spillover – an experiment in a college classroom* (Working Paper No. 12251). Cambridge, MA: National Bureau of Economic Research.

Harsanyi, J. C. (1962). Measurement of social power, opportunity costs, and the theory of two-person bargaining games. *Behavioral Science*, *7*, 67–80.

Holt, C., & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Howe, L. D., Hargreaves, J. R., & Huttly, S. R. (2008). Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. *Emerging Themes in Epidemiology*, *5*(3).

Huysentruyt, M., Barrett, C. B., & McPeak, J. G. (2009). Understanding declining mobility and inter-household transfers among East African pastoralists. *Economica*, *76*(302), 315–336.

Illife, J. (1987). *The African poor: A history*. Cambridge, UK: Cambridge University Press.

Jejeebhoy, S. (2002). Convergence and divergence in spouses' perspectives on women's autonomy in rural India. *Studies in Family Planning*, *33*(4), 299–308.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889.

Jolis, A., & Yunus, M. (2003). *Banker to the poor: micro-lending and the battle against world poverty*. New York: PublicAffairs.

Joshi, A., Lazarova, M. B., & Liao, H. (2009). Getting everyone on board: The role of inspirational leadership in geographically dispersed teams. *Organization Science*, *20*(1), 240–252.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change*, *30*(3), 435–464.

Karlan, D., & Zinman, J. (2011). Microcredit in theory and practice: using randomized credit scoring for impact evaluation. *Science*, *332*, 1278–1284.

Kenya National Bureau of Statistics. (2007). *Basic report on well-being: based on Kenya Integrated Household Budget Survey 2005 \06.* Nairobi, Kenya: Kenya National Bureau of Statistics.

Kenya National Bureau of Statistics and Society for International Development. (2013). *Exploring kenya's inequality: pulling apart or pooling together?* Nairobi, Kenya: Kenya National Bureau of Statistics.

Keswell, M., & Carter, M. R. (2014). Poverty and land redistribution. *Journal of Development Economics*, *110*, 250–261.

Kusago, T., & Barham, B. L. (2001). Preference heterogeneity, power, and intrahousehold decision-making in rural Malaysia. *World Development*, *29*(7), 1237–1256.

Lazear, E. P. (1999). Globalisation and the market for team-mates. *The Economic Journal*, *109*(454), 15–40.

Little, P. D., McPeak, J. G., Barrett, C. B., & Kristjanson, P. (2008). Challenging orthodoxies: understanding poverty in pastoral areas of East Africa. *Development and Change*, *39*(4), 587–611.

Lybbert, T. J., Barrett, C. B., Desta, S., & Coppock, D. L. (2004). Stochastic wealth dynamics and risk management among a poor population. *The Economic Journal*, *114*(498), 750–777.

Malhotra, A., Schuler, S., & Boender, C. (2002). *Measuring women's empowerment as a variable in international development* (Paper prepared for the World Bank Workshop on Poverty and Gender). Washington, DC: World Bank.

Matin, I., Sulaiman, M., & Rabbani, M. (2008). *Crafting a graduation pathway for the ultra poor: Lessons and evidence from a BRAC programme* (Working Paper No. 109). Dhaka, Bangladesh: BRAC Research and Evaluation Division.

McPeak, J. G. (2006). Confronting the risk of asset loss: What role do livestock transfers in northern Kenya play? *Journal of Development Economics*, *81*(2), 415–437.

McPeak, J. G., & Barrett, C. B. (2001). Differential risk exposure and stochastic poverty traps among East African pastoralists. *American Journal of Agricultural Economics*, *83*(3), 674–679.

Merttens, F., Hurrell, A., Marzi, M., Attah, R., Farhat, M., Kardan, A., & MacAuslan, I. (2013). *Kenya Hunger Safety Net Programme monitoring and evaluation component: impact evaluation final report 2009 to 2012.* Oxford, United Kingdom: Oxford Policy Management.

Nature Conservancy. (2013, July). *Kenya's national gender context and its implications for conservation: A gender analysis.* Nairobi, Kenya: Nature Conservancy, Central Science. Retrieved from http://www.nature.org/science-in-action/leading-with-science/kenya-gender-analysis.pdf

Nichols, A. (2007). Causal inference with observational data. *The Stata Journal*, *7*(4), 507–541.

Nichols, A. (2008). Erratum and discussion of propensity-score reweighting. *The Stata Journal*, *8*(4), 532–539.

Osterloh, S., & Barrett, C. B. (2007). The unfulfilled promise of microfinance in Kenya: the KDA experience. In C. B. Barrett, A. Mude, & J. Omiti (Eds.), *Decentralisation and the social economics of development: lessons from Kenya.* Wallingford, UK: CABI.

Ostrom, E. (2010). Beyond markets and states: polycentric governance of complex economic systems. *American Economic Review*, *100*(3), 641–672.

Page, S. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Parente, P. M., & Santos Silva, J. (2016). Quantile regression with clustered data. *Journal of Econometric Methods*, *5*(1), 1–15.

Peterman, A., Schwab, B., Roy, S., Hidrobo, M., & Gilligan, D. (2015). *Measuring women's decision making: Indicator choice and survey design experiments from cash and food transfer evaluations in Ecuador, Uganda, and Yemen* (Discussion Paper No. 01453). Washington, DC: IFPRI.

Putnam, R. (1993). *Making democracy work*. Princeton, NJ: Princeton University Press.

Russell, B. (1938). *Power: a new social analysis*. London: Allen & Unwin.

Santos, P., & Barrett, C. B. (2010). Identity, interest and information search in a dynamic rural economy. *World Development*, *38*(12), 1788–1796.

Santos, P., & Barrett, C. B. (2011). Persistent poverty and informal credit. *Journal of Development Economics*, *96*(2), 337–347.

Schatz, E., & Williams, J. (2012). Measuring gender and reproductive health in Africa using Demographic and Health Surveys: The need for mixed-methods research. *Culture, Health and Sexuality*, *14*(7), 811–826.

Schoar, A. (2009). *The divide between subsistence and transformational entrepreneurship.* Cambridge, MA: MIT.

Silvestri, S., Bryan, E., Ringler, C., Herrero, M., & Okoba, B. (2012). Climate change perception and adaptation of agro-pastoral communities in Kenya. *Regional Environmental Change*, *12*(4), 791–802.

Upadhyay, U., & Karasek, D. (2012). Women's empowerment and ideal family size: An examination of DHS empowerment measures in sub-Saharan Africa. *International perspectives on sexual and reproductive health*, *38*(2), 78–89.

Valdivia, M. (2015). Business training plus for female entrepreneurship? Short and medium-term experimental evidence from Peru. *Journal of Development Economics*, *113*, 33–51.

van den Bold, M., Quisumbing, A., & Gillespie, S. (2013). *Women's empowerment and nutrition: an evidence review* (Discussion Paper No. 01294). Washington, DC: IFPRI.

Warui, H. M., & Kshatriya, M. (2009). Implications of community based management of woody vegetation around sedentarised pastoral areas in the arid northern Kenya. *Field Actions Science Reports*, *3*.

World Bank. (2012). *World Development Report: Jobs*. Washington, DC: World Bank.