NBER WORKING PAPER SERIES

PSYCHOSOCIAL CONSTRAINTS, IMPACT HETEROGENEITY AND SPILLOVERS IN A MULTIFACETED GRADUATION PROGRAM IN KENYA

Geyi Zheng
Michael Carter
Nathan Jensen
Laurel Krovetz

Working Paper 31611
http://www.nber.org/papers/w31611

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023

This study was made possible through the generous support of the American people through the United States Agency for International Development Cooperative Agreement No. AID-OAA-L-12-00001 with the BASIS Markets, Risk and Resilience Feed the Future Innovation Lab. The contents are the responsibility of the authors and do not reflect the views of the US Government. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Psychosocial Constraints, Impact Heterogeneity and Spillovers in a Multifaceted Graduation Program in Kenya
Geyi Zheng, Michael Carter, Nathan Jensen, and Laurel Krovetz
NBER Working Paper No. 31611
August 2023
JEL No. D91,O15

ABSTRACT

Poverty reduction programs modeled on BRAC's graduation approach build up both tangible productive assets and intangible psychosocial assets such as self-confidence and the aspiration for upward mobility. The goal of this paper is to better understand how psychosocial factors operate and shape the impact of graduation programs. After deriving a set of hypotheses about the impacts of psychosocial constraints from a dynamic optimization model of the choice between a low income, casual wage-labor occupation and a higher earning entrepreneurial activity, this paper exploits a randomized controlled trial of a graduation program implemented in the pastoralist regions of Northern Kenya. Key empirical findings include that the estimated highly favorable average treatment effects disguise substantial heterogeneity, with beneficiaries who began with severe depressive symptoms gaining little from the program. The RCT's saturation design also allows us to identify substantial spillover effects onto the asset accumulation of women who were not enrolled in the graduation program. Spillovers are also estimated to positively affect non-beneficiary women's preference for upward economic mobility, providing a plausible explanation for their accumulation of capital despite no direct support from the graduation program. The paper draws out the implications of these findings for the cost-effective design and implementation of graduation programs.

Geyi Zheng
Department of Agricultural &
Resource Economics
Social Sciences & Humanities Building
1 Shields Avenue
Univeristy of California
Davis, CA 95616
gyzheng@ucdavis.edu

Michael Carter
Department of Agricultural and
Resource Economics
University of California, Davis
One Shields Avenue
Davis, CA 95616
and NBER
mrcarter@ucdavis.edu

Nathan Jensen
University of Edinburgh
njensen@ed.ac.uk

Laurel Krovetz
Warren Hall
Cornell University
Ithaca, NY 14853
lk436@cornell.edu

A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/5822

1 Introduction

Multifaceted "graduation" programs transfer tangible productive assets and also intensively mentor transfer recipients to build up their skills, self-confidence and aspirations.[1] Asset transfers are meant to relax capital constraints, whereas mentoring is meant to relax what Bossuroy et al. (2022) call psycho-social constraints to asset accumulation and the shift to more productive livelihoods and occupations. While capital constraints are relatively well-understood, our goal in this paper is to better understand psychosocial constraints, how they operate and how they shape the direct and spillover impacts of graduation programs.

We begin with a dynamic stochastic optimization model of the choice between a low income, casual wage-labor occupation and a higher earning but risky business activity that relies on capital and entrepreneurial acumen. This model elucidates the potential sources of the pronounced heterogeneity in impacts that has been empirically observed in evaluations of graduation programs (e.g., Bandiera et al. (2017) and Gobin et al. (2017)). When combined with insights from the economics of depression (de Quidt and Haushofer (2016)), the theory suggests an empirical strategy for identifying one important source of that heterogeneity, namely the severity of initial depressive symptoms in the beneficiary population. In addition, we also use the theoretical occupational choice model to illustrate how a graduation program might be expected to relax psychosocial constraints through an endogenous preference mechanism that operates through the "sour grapes" or adaptive preferences described by Elster (1983) in which preferences adapt to constraints. Importantly, the theory illustrates how this endogenous preference mechanism would be expected to generate social spillovers.

To examine the empirical veracity of these ideas, we implemented a randomized controlled trial of a woman-targeted graduation program implemented by the BOMA Project NGO in the pastoral regions of Northern Kenya. We first show that conventionally measured average treatment effects (ignoring spillovers) are sizable, as treated women's holdings of productive assets, family cash income and financial savings all significantly increase by 414%, 12% and 500% 24 months after the program began. However, hiding beneath these

[1] This approach was pioneered by the NGO BRAC which had noticed that their signature micro-credit programs were not appropriate for the poorest. BRAC thus developed their "Targeting the Ultra-Poor," or TUP program that intended to build the tangible and intangible assets of the poorest such that they were ready to graduate to micro-credit (see the discussion in Hulme and Moore (2008) and Hashemi (2011)).

average treatment effects is substantial impact heterogeneity. First using an a-theoretic conditional quantile analysis, we show that approximately 25% of the beneficiary population experienced no benefits form the program, a finding similar to those in the Bandiera et al. (2017) and Gobin et al. (2017) studies. Digging deeper, and following the lead of our theoretical modeling, we show that the almost 20% of the beneficiary population that exhibited severe depressive symptoms at baseline had drawn down about half the assets transferred to them, had experienced no income gain, but had built up savings stocks at a rate similar to that of non-depressed women.

To explore endogenous preference effects and social spillovers, we leverage the rollout and saturation design of the RCT which allows us to define an exposure measure to beneficiaries of the BOMA graduation program. Defined as the probability that a random social interaction would be with a BOMA project beneficiary during the 24-month period between the baseline and follow-up, this measure varies between 0 and 60%. Using this measure reveals statistically significant spillovers on both beneficiary and non-beneficiary women.[2] For non-beneficiary women, the impacts suggest an asset accumulation rate at 25% of that for beneficiary women, even though the former received no direct asset grant or other support from the BOMA program. Employing a ladder-of-life-based measure of the desire for upward economic mobility, we show that a plausible explanation for these real spillovers is a statistically significant impact of exposure to BOMA beneficiaries on the non-beneficiaries' preference for upward mobility.

The remainder of the paper is organized as follows: Section 2 presents the dynamic optimization model of occupational choice and highlights its key implications regarding psychosocial constraints. Section 3 describes the BOMA Project graduation program and the research design employed to assess its direct and spillover impacts. Section 4 estimates average treatment effects and impact heterogeneity related to baseline mental health indicators. Section 5 introduces the spillover exposure measure and estimates its effects on economic variables and preferences. Finally, Section 6 concludes by discussing the implications of these findings for designing more cost-effective graduation programs.

[2] Taking these spillovers into consideration and calculating the total causal effect of the BOMA program raises the benefit-cost ratio by 29%.

2 Theoretical Perspectives on Psychological Assets and Graduation Programs

Poverty trap models that allow exploration of non-linear asset dynamics offer a compelling framework to understand the situations of individuals targeted by graduation programs who seem trapped in chronic poverty. This section begins by presenting a dynamic stochastic model of occupational choice. It is demonstrated that this model predicts systematic heterogeneity in the outcomes of graduation programs, which can potentially be identified. Furthermore, the latter part of this section illustrates how endogenous "adaptive" preferences would be expected to operate and generate spillover effects from treated to non-treated populations.

2.1 A Dynamic Theory of Occupational Choice and the Heterogeneous Impact of Graduation Programs

Consider an economy comprised of individuals each endowed with an initial level of wealth (k_{j0}) and a level of entrepreneurial skill (α_j), as suggested by Buera et al. (2014). Note that for individuals who have never been entrepreneurs, α_j is a latent variable. In this model, individuals can devote their resources to one of two different occupations:

- *Casual Wage Labor* which generates income $F^w_{jt} = w_0 + f^w(k_{jt})$; or,
- *Entrepreneurial Occupation* which generates income $F^e_{jt} = (w_0 - A) + \alpha_j f^e(k_{jt})$.

Note that w_0 is the returns from full-time work in the causal labor market. Income can also be earned from accumulated capital wealth, through either a low-return investment associated with the wage labor occupation (F^w), or a higher returning entrepreneurial investment (F^e) that requires withdrawal of a discrete amount of time from the labor market (to run the business), $\frac{A}{w_0}$, and is sensitive to the agent's level of entrepreneurial skill. We assume that both investment functions are increasing and concave in k,[3] that $f_e(k) > f_w(k) \forall k$ and that $0 \leq A \leq w_0$. Combining these two income generation processes yields a a non-concave

[3] This assumption of course allows f^w to be linear in k, as it would be if capital under the casual wage labor were put into a simple savings, or buried in the backyard, earning no interest.

income possibility set with locally increasing returns to scale: $F(\alpha, k) = max[F^w, F^e]$.[4] To ease discussion, we assume that earnings from casual wage labor occupation are below the poverty line and that someone who persists in that occupation is chronically poor.

Following Ikegami et al. (2019), we assume that capital wealth is subject to shocks and evolved according to:

$$k_{jt+1} = (k_{jt} + f(k_{jt}) - c_{jt})(\theta_{jt+1} - \delta)$$

where c_{jt} is consumption, $0 \leq \theta_t \leq 1$ is a random capital depreciation shock with known probability distribution function and δ is a standard, fixed rate of capital depreciation.[5]

To study the dynamics of occupational choice and consumption dynamics, we assume that individuals solve the following inter-temporal maximization problem:

$$\max_{c_{jt}} \quad E_\theta \sum_{t=0}^{\infty} \beta^t u(c_{jt})$$

subject to:

$$\begin{aligned} c_{jt} &\leq k_{jt} + F(\alpha_j, k_{jt}) \\ F(\alpha_j, k_{jt}) &= max[F^w, F^e] \\ k_{jt+1} &= (k_{jt} + f(k_{jt}) - c_{jt})(\theta_{jt+1} - \delta) \\ k_{jt} &\geq 0 \end{aligned} \tag{1}$$

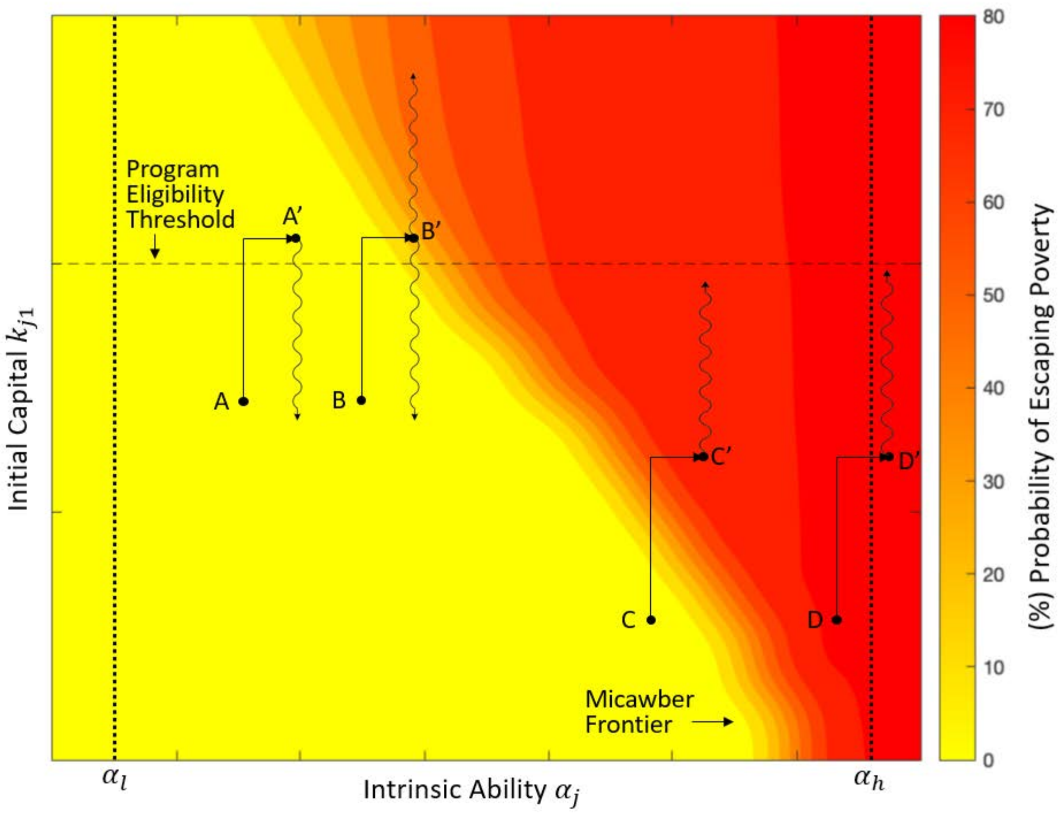
where E_θ is the expectation taken over the distribution of the negative shocks and β is the time discount factor. $u(c_{jt})$ is the utility function defined over consumption.

The key question that this model allows us to address is from which positions in the α, k space is it optimal to accumulate capital and adopt the higher income, entrepreneurial occupation. Figure 1, taken from Ikegami et al. (2019), provides the answer to that question based on numerical analysis of dynamic optimization problem 1. The horizontal axis measures the entrepreneurial skill parameter, α, while the vertical axis measures initial tangible

[4] Note that even an individual trying to join the entrepreneurial class will operate the wage labor process until she has sufficient capital to make it worth while to begin to run the entrepreneurial technology.

[5] Under this specification, both terms of capital are subject to depreciation through theft or other mechanism of loss.

Figure 1: Optimal Occupational Choice & The Heterogeneous Impact of Asset Building



productive capital, k. For purposes of this graph, we assume that decision makers know their true skill parameter. The probabilities associated with the heat map (and captured in the side bar in the figure) measure the chances that the individual will, in the long-run, end up in the high earning, entrepreneurial equilibrium.[6]

As can be seen, the probability of escaping the low earning, casual labor occupation (and chronic poverty) is 0 for initial positions on the west side, or yellow-colored portion of the diagram. The boundary between this portion of the diagram and the portion where

[6]To derive these probabilities, we numerically solve the model for a wide array of initial asset positions over a number of randomly drawn shock sequences. Specifically, for each of 1500 initial positions evenly distributed across the initial endowment space shown in Figure 1. The infinite horizon model was solved for each asset position, generating an optimal consumption value as well an optimal asset holding. A random shock was then generated, assets were updated and infinite horizon model was again solved for each updated asset position. This procedure was repeated 60 times, yielding a single history of consumption, income and assets for each initial asset position. At the end of each 60-year, an indicator variable was formed indicating whether or not the individual was pursuing the wage labor or the entrepreneurial livelihood in period 60.

This entire process was then repeated 1000 times, generating 1000 histories for each of the 1500 initial endowment positions. The heat map in Figure 1 displays the probability that an individual at the indicated initial asset position will end up at the higher income entrepreneurial occupation across the 1000 histories.

that probability falls becomes greater than 0, demarcates what Ikegami et al. (2019) call the Micawber Frontier. From asset positions west/south-west of that boundary, it is not dynamically optimal in the sense of optimization problem 1 to even attempt to become an entrepreneur. To the north/northeast of that boundary, it is optimal to try to accumulate and advance to the higher income entrepreneurial equilibrium. However, as can be seen, not everyone who attempts to advance will succeed in the stochastic environment captured in the optimization problem as the probabilities of escaping poverty just to the right of the frontier between 1 and 0.

As can also be gleaned from the figure, there is a critical minimum value of α_ℓ below which it is not optimal to stay at the high equilibrium even if the individual were gifted a large capital stock. Foreshadowing later discussion, this would be true even for an individual who incorrectly perceives her entrepreneurial capabilities to be below α_ℓ. Note also, that individuals with values of $\alpha > \alpha_h$ will (almost) always escape chronic poverty and move to the high income entrepreneurial equilibrium.

2.2 Implications of Dynamic Model for Graduation Program Impact Heterogeneity

We can use Figure 1 to derive predictions about the impact of a graduation program like that implemented by the BOMA Project in Northern Kenya. By design, graduation programs are means-tested based on observable characteristics. For illustrative purposes, we assume that any individual observed with capital below the dashed horizontal line is eligible for the graduation program, whereas those above it are not. Under the numerical parameterization used to derive Figure 1, this horizontal line is approximately mid-way between the low and high steady state capital stocks. As discussed earlier, multi-faceted graduation programs transfer tangible assets, moving beneficiaries to the north in the figure. They also transfer intangible assets including hard entrepreneurial skills, as well as soft skills intended to boost beneficiaries confidence in their own capabilities. This second transfer is portrayed as a rightward shift in Figure 1.

We have projected onto the figure four types of asset positions, labelled A to D. We illustrate the first order impact of the graduation program as a shift to the northeast, moving a beneficiary at A to A', B to B', *etc.* We can use the probabilities shown on the graph to predict the expected outcome of the graduation program on each of these stylized positions. If we assume that individuals are distributed across the entire domain of α values shown in the Figure, then we can immediately appreciate the heterogenous impacts that this stylized, multi-faceted graduation program will have. Four kinds of impacts are possible depending on the individual position in the tangible-intangible asset space (impacts means impact relative to an otherwise identical individual who did not receive and was uninfluenced by the program):

1. *Position A:* The asset building transfers move the individual to point A', but are inadequate to move the individual above the Micawber frontier. The optimization problem above implies that the optimal policy for the individual is to draw down (consume) their tangible assets such that in the long-run the individual ends up at the low level equilibrium. Were we to compare that individual to an otherwise identical person who did not receive the transfers, we might observe short-term effects on consumption and income, but in the long run the impact would be zero.

2. *Position B:* The Asset transfers move the individual into the multi-colored band, where she may or may not succeed in reaching the high equilibrium (note that B' is in a position where the chances of escaping the low equilibrium is only 25%). Average causal impacts of a graduation program for individuals in this position would be a mix of those who escaped the low level equilibrium and those who did not.

3. *Position C:* Graduation program asset transfers would move this person to position C', which is firmly in the zone where the probability of reaching the entrepreneurial equilibrium in the long term poverty is quite high. Impacts for individuals in this position (compared to someone at C who did not receive the program) would be expected to be substantial over the long-term as beneficiaries would likely move from the low to the high equilibrium.

4. *Position D:* Individuals beginning at point D would also be predicted to move the high equilibrium in the long run. While graduation program transfers would help

them approach the high equilibrium more quickly than a control person at position D, the longer term impact would be predicted to be zero as the control person is also predicted to escape poverty and also end up at a similar long-run position

As this overview of the possible impacts of a graduation program makes clear, any average treatment effect estimated from a well-balanced sample would be expected to be a data-weighted average of these different, and highly heterogenous, effects.

2.3 Economics of Depression and the Heterogeneous Impact of Graduation Programs

As is apparent from Figure 1, the expected impact of a graduation program will depend fundamentally on an individual's initial wealth and their entrepreneurial acumen. While conditional quantile analysis can reveal some information about impact heterogeneity (see Section 4.2), one barrier to a more structural analysis of heterogeneity stems from the fact that entrepreneurial skill (α) is not directly observable. Importantly, entrepreneurial skill is not only unknown to the econometrician, but it is also unknown to most chronically poor women who do not have lived experience as entrepreneurs. While we might imagine that in the absence of empirical evidence, most women would hypothesize themselves to have an average level of α, there is one, observationally distinguishable group that we would expect to systematically understate their true entrepreneurial skill.

Citing psychiatrist Aaron Beck's exposition on depression (Beck, 1967), de Quidt and Haushofer (2016, 2019) note that depression leads individuals to underestimate their intrinsic abilities. Depression, considered one of the most common mental illnesses, specifically major depressive disorder, encompasses a constellation of disruptive symptoms affecting mental well-being. This well-being is defined as 'a state in which the individual realizes their abilities, can cope with life's normal stresses, remains productive, contributes to their community' (World Heath Organization, 2017). Formally modeled by these authors, depression prompts individuals to behave as if their entrepreneurial efficacy or skill (analogous to α in our model) is lower than its actual level. Given that impoverished women trapped in the low equilibrium lack personal experiences to counterbalance this underestimation, we can hypothesize that depression influences decisions as if individuals are situated on the left side of Figure 1. Even assuming that there is no intrinsic correlation between entrepreneurial skill and depression, this perspective suggests that women with depressive symptoms will disproportionately act as if they are low skill types.[7] Section 5.3 will return to use this insight to see if lower graduation program impacts are indeed found amongst those with high baseline levels of depressive symptoms.

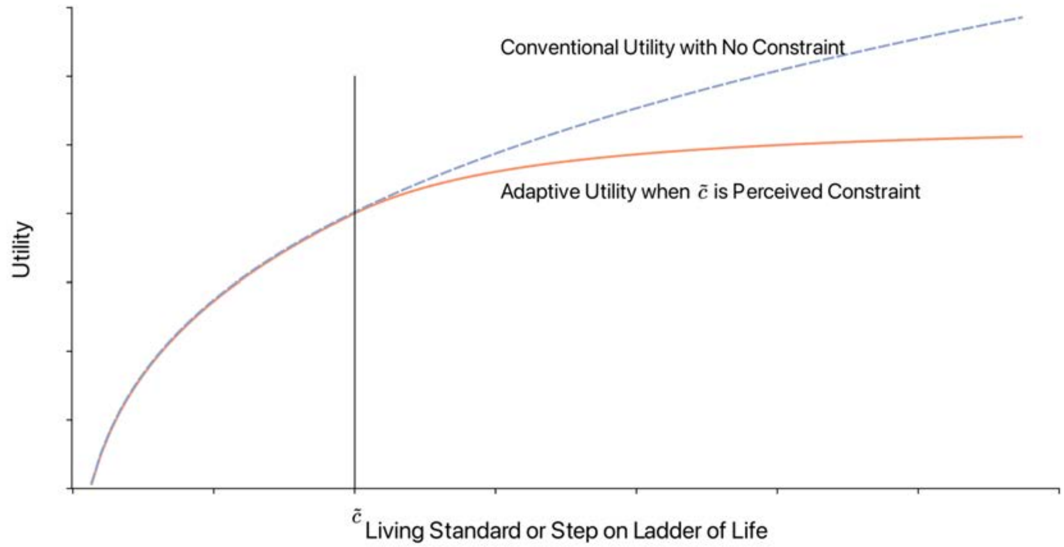
2.4 Adaptive Preferences and the Psychosocial Spillovers of Graduation Programs

While graduation programs could generate an array of spillover benefits (see Section 5 below), this section lays the conceptual foundation for a specific type of spillover that could be particularly relevant in our study area, namely one that operates through an endogenous preference mechanism. Specifically, we build on the notion of adaptive preferences put forward by Elster (1983). Elster argues that although conventional economic analysis views the preferences that guide utility maximization as independent from the constraints that limit the maximum achievable utility, the latter in fact can reshape the former. To illustrate the notion that preferences adapt to constraints, Elster draws on the Aesop's fable, "Sour Grapes." In the fable, a fox wants to enjoy a bunch of grapes hanging from a vine. Despite multiple attempts to reach the desired grapes, the fox can never reach the not quite low-hanging fruit and ultimately walks away saying he did not really want the grapes after all because they were sour. In the fable, constraints cause preferences to change.

One approach to encapsulate Elster's concepts is by proposing that we deduce our potential highest achievable living standard by observing individuals whom we perceive as similar to ourselves. This aligns with Appadurai's notion of an 'aspiration window,' which individuals use to assess whether others are sufficiently akin to them to serve as credible models for adopting reasonable levels of living standards (Appadurai, 2004). For a person i we write this perceived maximum attainable living standard as $\tilde{c}(c_{g(i)})$, where $c_{g(i)}$ is the vector of

[7] de Quidt and Haushofer (2016) also note that depression can lower effective labor supply. Incorporating this insight into the model above as an decrease in w_0, will result in a further shift right in the Micawber threshold, again making it more likely that depressed people will find themselves in a position like A in Figure 1.

Figure 2: Adaptive, Sour Grapes Preferences



observed living standards of those in i's social reference group $g(i)$.[8] The notion of adaptive preferences is that we sour on living standards that we see as unattainable, undervaluing living standard advances beyond $\tilde{c}$ relative to what we value them if we thought they were attainable. In other words, adaptive preferences suggest an endogenous consumption threshold beyond which the utility function flattens out, indicating that we expect few gains in our subjective well-being from material advance beyond the threshold level:

$$u(c_{it}) = \begin{cases} u^{\ell}(c_{it})\ if\ c_{it} < \tilde{c}(c_{g(i)}) \\ u^{h}(c_{it})\ otherwise \end{cases} .$$

Figure 2 illustrates this preference structure. The upper, dashed-blue curve is a standard constant relative risk aversion utility function. The lower solid, orange curve represents an adaptive preference function for an individual who sees her maximum attainable living standard as $\tilde{c}$. Adapting to that constraint, the individual ascribes little incremental value to advance beyond $\tilde{c}$ as the utility function flattens out beyond that level.

The notion of a threshold consumption level where preferences discontinuously change has featured in a number of models of endogenous preferences (*e.g.*, Genicot and Ray (2017),

[8]For example, $\tilde{c}$ could be the mean, median or max of the vector $c_{g(i)}$.

Lybbert and Wydick (2018)). Adaptive preferences as presented here is built on the notion that we are happier (less frustrated) when we do not highly value unattainable living standards and instead are happy enough with the attainable. Elster suggests that this tendency to devalorize what we see as unobtainable is consistent with cognitive dissonance theory.[9]

Section 5.3 below presents a strategy for empirically measuring endogenous adaptive preferences. We close this section by using the dynamic occupational choice model (1) to illustrate the behavioral implications of a shift in endogenous adaptive preferences. The probabilities for upward mobility illustrated in Figure 1 were derived assuming adaptive preferences with a low critical value, set in the vicinity of the steady state income for someone pursuing the poor, informal wage-labor occupation. We then ask what happens if the threshold level shifts up to the steady state level of the entrepreneurial occupation, perhaps because members of the community begin to move forward economically. Figure 3 shows the results of resolving the dynamic model using these new preferences. The color scale shows the change in the probability of reaching the entrepreneurial equilibrium after the preference shift. The black solid curve projected across the figure is the Micwaber Frontier from Figure 1. As can be seen, the largest change in the probability of economic advance are for those endowment positions that are just to the west of the original Micawber Frontier. A shift in preferences fundamentally alters behavior for individuals at these endowment positions, indicating that the preference change would induce them to try to accumulate productive assets and reach the higher income entrepreneurial equilibrium Note also that the preference change has no effect on the behavior of those far away from the original Micawber Frontier.

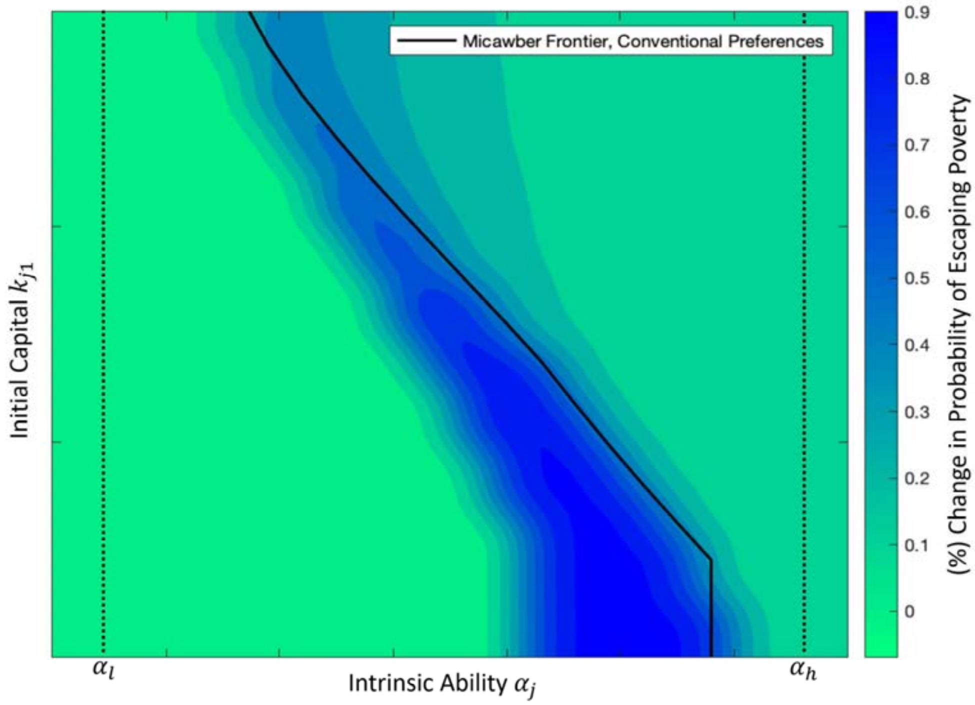
3 Program Intervention and Research Design

This study examines the impacts of the REAP program on its participants and those in their community. It is part of a larger agenda that is assessing not only the impacts of REAP, but also the impacts of the Index Based Livestock Insurance (IBLI)[10] product, and

[9]Economic analyses that draw on cognitive dissonance theory include Montgomery (1994) and Laajaj (2017).

[10]IBLI is a commercial insurance product that households can purchase from Takaful Insurance of Africa. IBLI policies last 12 months and make payouts if a proxy for forage conditions in the insured area fall below a threshold. The proxy is based on the Normalized Differenced Vegetation Index (NDVI) collected

Figure 3: Endogenous Preferences & Change in the Prospects of Upward Mobility

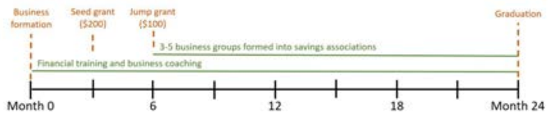


will test for synergies between the two. The larger agenda included randomized treatments of both REAP and IBLI and has three waves of data collection: 2018, 2020, and 2022. This manuscript focuses on the heterogeneity of REAP's impacts and its spillovers on others. As will be discussed in the following section, the research design was developed specifically to examine REAP's spillovers, but does so using only the 2018 and 2020 data. At the same time, the assessment of IBLI requires data from the 2022 collection, which is after several droughts had moved through the region. For this reason, we leave the discussion of IBLI and related treatments out of this manuscript, except for those cases in which it is relevant for this study.

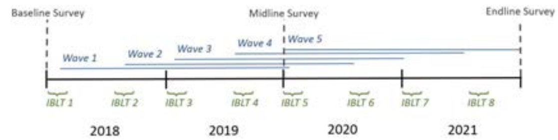
by NASA through sensors located on satellites and processed by the USGS. More on IBLI can be found at https://ibli.ilri.org/. For more on NDVI, visit https://www.usgs.gov/landsat-missions/landsat-normalized-difference-vegetation-index.

Figure 4: Intervention and Study Design

(a) Timeline for REAP Intervention



(b) Study Timeline



3.1 The REAP Graduation Program

The REAP program aims to help poor women move out of, and stay out of, poverty by providing financial and business training, and mentoring participants as they develop and launch a local business. The REAP intervention targets the poorest women using a community-based assessment procedure known as the Participatory Rural Appraisal (PRA). [11] Once identified through the PRA, BOMA staff then visit each of the PRA-qualified households to verify that they are consistent with their PRA classification and to determine if they have a member that is eligible for the REAP program. Eligibility requires that the participant is female, is of productive age, is of sound mind, is a permanent resident of the community, and does not suffer from drug addiction. [12]

Once enrolled, participants of the REAP program receive mentoring and skills training from a BOMA staff member throughout the duration of the program. BOMA staff also help participants form into 3-person business groups and to develop a business plan. Upon approval by the mentor, the 3-person business receives a seed grant worth approximately USD200 with which to purchase any needed assets and the starting inventory for the business. A "jump grant" worth approximately USD100 is provided to the business after three months of operation as long as they have been adhering to the principles of the program. The left panel of Figure 4 coarsely illustrates the timelines of the REAP intervention.

BOMA relies on full-time staff, called mentors, to implement the REAP program. Each mentor is assigned a catchment area and is responsible for the training and monitoring of

[11] The PRAs involve working with community members to identify locally defined wealth groups and then allocating community members into them. It is common for communities to agree on four to five wealth categories and allocate about half of the households into the lowest two categories. No matter the number of wealth categories that the communities uses, BOMA targets women from the lowest two.

[12] Reportedly, alcoholism is common disease in some regions that BOMA operates in and can cause considerable dysfunction among the business members.

participants in their area. For practical reasons, the mentors enroll and implement the REAP intervention in waves, so that at any point a mentor might be working with women that are at several different stages of the REAP program in the same community. These waves are commonly rolled out every six months.

3.2 Research Design

This research was implemented in northern Samburu County, Kenya. This study region was selected because the BOMA Project had plans to launch REAP there (relevance) but had not yet done so (uncontaminated). The pastoral and agro-pastoral communities of northern Samburu County are similar to the communities that BOMA works with across northern Kenya; they are remote, suffer from high levels of poverty, and they are often unable to take advantage of and, or are overlooked by government and development interventions. Most households live in relatively small clustered settlements called manayattas. In the remainder of this paper, we will simply refer to manyattas as communities.

The study's research design and sample selection was performed to ensure comparability between the treatment and control groups. To begin with, BOMA completed the PRA and verification process, and then provided the research team with a roster of REAP-eligible women. The research team then stratified the roster by community, and used a randomization process to distribute participants into those that would participate in the study and those that would enter the pool of REAP-eligible, but non-study, individuals. Within each community, those that were selected to participate in the study were provided with a random rank. When it was time to enroll new participants in REAP, those with the lowest rank were offered the opportunity to enroll in REAP. If they excepted, they were then designated "anchor women". Each anchor woman would then select two women from the pool of REAP-eligible non-study individuals to participate in their 3-person REAP business. This approach allowed us to maintain an uncontaminated and relevant control group, while also maximizing the power of the sample size by splitting all the REAP-treated study participants into different REAP businesses. Further, it provided a clear and exogenous process for identifying study participants that would be recruited for the next wave of REAP in a context in which we expected reasonably high attrition.[13]

In total, 88 communities participated in the project at baseline and the number of control and anchor women selected in each community was set in proportion to the number of REAP-eligible women in each community. A sample of 1,502 REAP-eligible women enrolled in the study, and the study had a target of 700 anchor women that would participate in the REAP program. The study also included a sample of 373 "vulnerable" women from the same 88 communities that were drawn from the population identified in the PRAs as being from the wealth category just above the REAP-eligible threshold. We call these women "vulnerable" because they are close to, but not in, the REAP-eligible (poorest) categories. These vulnerable women were not eligible to receive the REAP treatment but could be impacted by spillovers from it and were eligible to receive the IBLI treatment. As Section 5.1 discusses in detail, the RCT was designed to detect spillovers on to all sub-populations.

3.3 Integrity of the experimental design

Table A2 in Appendix B present the baseline test for the outcome variables used as the primary outcome measures in the study. Panel A presents the mean comparisons and t-tests for equality of means between the treatment and control groups of poor households in the sample. The treatment status is defined based on assignment to treatment by the endline survey. We observe no significant difference in either the primary outcome variables or the household characteristics at the baseline. While the difference of means of treatment and control poor households on household earnings is significant at the 10% level, the aggregate test, reported in Panel B of Table A2, finds that we are not able to reject the equality of means across all the measures (p-value = 0.11). Overall, the sample balance was good between the treatment and control poor groups.

Table A3 in Appendix B presents an analysis of survey attrition for both midline and endline data collections. The follow-up rate was excellent. We managed to survey 92% of

[13] During the study design phase of the project, we assumed that attrition rates between the 2018 and 2020 study would be a relatively high rate (10%) in part because the population is migratory by nature and we planned to drop households that moved out of the study region and becuase there would be a large number of study participants–both those that would be control and those that would not enroll in REAP unitl later waves–that were not recieving any engagements from the project at all.

baseline respondents in the midline, and 86% in the endline (Panel A). We do not observe any significant differences in attrition rates across treatment and control groups. Panel B of Table A3 presents an analysis of the characteristics of respondents who were more likely to be resurveyed during the midline and endline. Panel C presents a test of whether being assigned to receive treatment affected the type of person who completed both midline and endline surveys. We did not find evidence that the treatment caused a sample composition bias by affecting attrition rates. We fail to reject that the treatment status indicator and all the outcome variables interacted with treatment status were zero. The p-values for the test are 0.66 (midline) and 0.17 (endline), thus supporting the contention that survey attrition did not lead to a different sample frame across treatment and control groups.

4 Average Treatment Effects and Impact Heterogeneity by Baseline Depressive Symptoms

This section presents our empirical analysis using the standard approach commonly found in the literature, which does not account for spillover effects. However, Section 5 will delve into the consideration of spillovers. In this section, our focus is solely on significant material outcomes, namely women's business assets, family cash income, and women's savings. Regarding women's business assets, we acknowledge that some participants only reported their private business assets instead of collective assets. To address this, we employed a conservative measurement for imputing the missing data on business assets. For detailed information on the imputation procedures, please refer to Appendix C. Results for several key psychological outcomes are presented in Appendix Table A6. Despite the program's intensive mentoring, these results indicate no impact on depressive symptoms, as measured by the Center for Epidemiological scale. However, there was an observed increase in individuals' perceived control over events that influence their lives, indicating an internal locus of control, as defined by Rotter (1966).

4.1 Conventional Average Treatment Effects

We start our discussion first by presenting standard intent-to-treat (ITT) treatment results, estimating the following ANCOVA model:

$$(2) \quad y_{hm} = \alpha_0 + \alpha_1 y^0_{hm} + \beta^a W^a_{hm} + \beta^b W^b_{hm} + \varepsilon_{hm}$$

where y_{hm} is the 2020 outcome variable of interest for individual h in community m, y^0_{hm} is the 2018 baseline value of that same variable, W^a_{hm} and W^b_{hm} are binary indicator variables for assignment to waves 1 or 2 and 3 or 4, respectively of the BOMA program.[14] The error term ε_{hm} is clustered by community. Under this specification, the control is comprised of women selected for eligibility for REAP, but not assigned to any of the first four treatment waves. A fraction of these women were later assigned to later treatment waves, but they did not change their status at the time of the 2020 follow-up survey.

The β^w $(w = a, b)$ parameters identify the intent-to-treat impact of the program under two assumptions: random assignment to treatment and no spillovers between treatment and control households. Later, we will relax the latter assumption by exploiting our saturation design. We can compare the treatment effects obtained from this initial estimation with the treatment effects that take into consideration the spillover effects on the control women within the same cluster.

The first and last columns of Table 1 report the OLS estimates of the β^w parameters in equation 2. As can be seen, the REAP program on the key economic outcomes. For waves 1 and 2 women, who on average have been enrolled in the program for 20 months by the time of 2020 data collection, women's total business assets increased by $PPP 190, household's annual cash income by $PPP 98 and amount deposited to savings increased by $PPP 56. All these estimated impacts are significant at the 1% level and represent increases of 414%, 12% and 500% over the 2018 baseline levels.[15]

We find no significant impact on household earnings for women who enrolled in waves 3

[14] After estimating the model with separate parameters for each of the four waves, we combined them into two groups as the coefficients on waves 1 and 2 were similar, as were those for waves 3 and 4.

[15] Hinting at some spillover effects, these percent increases are only 251%, 18% and 325% over control group 2020 levels.

Table 1: Average and Conditional Quantile Treatment Effects

	Treatment Waves 1-2					**Treatment Waves 3-4**
	Average Impact (OLS)	Conditional Quantile Estimates *Q25*	*Q50*	*Q75*	*q90*	*Average Impact (OLS)*
Women's Business Assets ($PPP)	190***	NE	161***	346***	534.9***	125***
	(22.3)	NE	(13.76)	(34.57)	(54.33)	(18.6)
Household Income ($PPP)	98***	60.06*	74*	124.7	203.4	4.4
	(34.2)	(20.75)	(37.24)	(106.5)	(129.8)	(36.3)
Women's Savings ($PPP)	56***	NE	NE	87***	189.9***	25***
	(8.27)	NE	NE	(12.69)	(24.54)	(6.2)
Observations				1385		

Notes: Regressions include baseline levels of the dependent variable. Standard errors for the average treatment effects are clustered at the community level. For the quantile regressions, standard errors were calculated using the bootstrap method with 20 replications. For some lower quantiles, impacts were not estimable (NE) because there was no variation in the dependent variable. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

and 4 (who on average had been in the REAP program for 9 months) but their reported total business assets increased by $PPP 125 and the amount deposited to savings also increased by $PPP 24.8. For women who have enrolled in the program for 18-24 months, we document some increase in their business assets. Moreover, we find evidence on transfer of business assets of the women group businesses into their household earnings and savings.

The REAP program, similar to other graduation programs, involves providing cash grants to women. By the time of the midline assessment, women enrolled in waves 1 and 2 would have received both a jump grant (amounting to $PPP 163 as an individual share from the three-woman business group) and a progress grant (amounting to $PPP 82 as an individual share from the three-woman business group if the business survives after 6 months). When comparing the treatment effects on women from waves 1 and 2 with the amount transferred through the program, we observe a decrease in the business assets held by the women's business groups. However, there is an increase in the reported cash income of these women's households and the amount deposited into savings.

For women who enrolled in the program during waves 3 and 4, we do not observe a significant increase in their household cash income, in contrast to the cohorts who started the program earlier. Additionally, the amount deposited into savings by these women is only 44% of what women from waves 1 and 2 have. These findings, which indicate that the assets accumulated by waves 3 and 4 women, is not surprising as many of these women were just starting their businesses at the time of the midline survey.

4.2 Conditional Quantile Analysis

We now present the Conditional Quantile Treatment Effects (QTEs) to emphasize the substantial heterogeneity in treatment effects that will be further explored in subsequent sections. As elucidated in the theoretical framework section, individuals display considerable variation in their initial ability levels. In the absence of engagement in entrepreneurial activities, these individuals frequently lack opportunities to realize their full potential. This diversity can result in highly heterogeneous treatment effects of the REAP program, which may not be fully captured by the overall Intent-to-Treat (ITT) effects.

By examining the conditional quantile function of economic outcomes in relation to the treatment indicators, we can evaluate whether the dispersion in earnings, total business assets, and savings increases or decreases with the REAP program. It's important to note that while the treatment effects on the percentiles we estimate here represent the residual distribution conditional on baseline levels and treatment wave indicators, our aim is to illustrate the diverse impacts that can occur within the targeted population. As demonstrated in the remaining section of Table 1, the treatment effect on the 90th percentile of the conditional distribution of total business assets is approximately three times that of the 50th percentile. Similarly, the disparities in household income between the 90th percentile and the 25th percentile of the conditional distribution are roughly 3.5 times the level observed at the 25th percentile.

For instance, among women falling within the 75th percentile and above in the total business assets distribution, provided their assignment to waves 1 and 2 and conditioned on the baseline business assets level, they begin accumulating greater business assets (1.4 - 2 times the initially transferred grants). This insight illuminates the diverse impacts of the program within the ultra-poor population, which is crucial for comprehending how to enhance the program's cost-effectiveness. In the subsequent sections, we will delve into exploring the observed heterogeneity in treatment impacts by examining the psychological pathways available for testing.

4.3 Unpacking Impact Heterogeneity by Baseline Depression

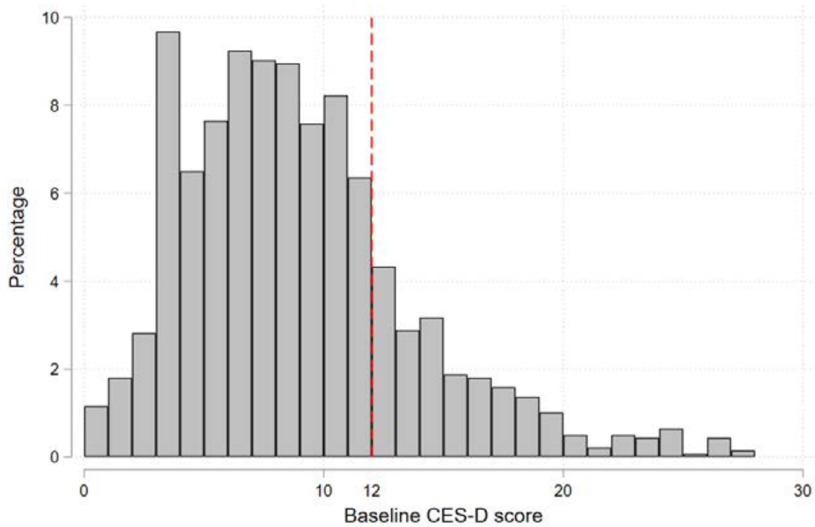
Using data from both the baseline and midline assessments, our intention is to observe how the treatment effects of REAP vary based on the respondents' baseline psychological assets, specifically their level of depression. Depression is measured in terms of the 10-point Center for Epidemiological Studies Depression (CES-D) score, which quantifies depressive symptoms. The CES-D (Radloff, 1977) score is a widely used measure in studies focusing on the impact of economic interventions or shocks on mental health. For instance, Christian et al. (2019) employs CES-D scores from waves 4 and 5 of the Indonesian Family Life Survey (IFLS) conducted in 2007 and 2014 as the depression measure.

In our study, CES-D scores range from 0 to 30, with higher scores indicating more severe depressive symptoms. A CES-D score of 12 is employed as a threshold to indicate clinically diagnostic depression. Notably, Baron et al. (2017) discuss studies conducted in the US and China that identify depression cut-off scores ranging from 8 to 16. In their own study in South Africa, the authors establish cut-off scores of 11, 12, and 13 based on the specific population. Hence, the depression cut-off score likely varies due to geographic, cultural, and population differences.

InKilburn et al. (2018), it is revealed that 37% of the youth (aged between 15 and 25) in their Kenyan sample had a CES-D score of 10 or higher, with a mean CES-D score of 8.61. In our sample of the poor group, 35.6% of women had a CES-D score of 10 or higher, and the mean score is 8.41. We adopt a threshold of 12 to indicate severe depression. Figure (5) displays the histogram of the baseline CES-D scores for the respondents, demonstrating that just under 20% of the poor group had a CES-D score exceeding 12.

As discussed earlier, depression might create an added vulnerability for women in poverty that will limit their ability to benefit from REAP. The model presented in Section 2 suggests that those not suffering from depression prior to treatment should be more likely to follow an upward trajectory towards the higher equilibrium outcome in the long run following treatment. On the other hand, the group considered depressed at baseline should be less likely to be placed on an upward trajectory, despite the program. The concept of depression is multifaceted and extends beyond the scope of underestimating one's own abilities (α in

Figure 5: Histogram of baseline CES-D scores



the Section 2 model). There may be other models that would link depression to the impact of a graduation program. Because we do not measure actual and perceived entrepreneurial ability, we cannot definitively identify the linkage posited by our theory.

To study the impact of depression on graduation program impacts, we modify regression equation 2 by adding an interaction between treatment and baseline depression dummy variable (D_{hm}) to identify the impacts on two sub-populations, baseline depressed and non-depressed groups:

$$
\begin{aligned}
y_{hm} = & \alpha_0 + \alpha_1 y^0_{hm} + \beta^a W^a_{hm} + \beta^b W^b_{hm} + \\
& D_{hm} \times \left[\delta^c + \delta^a W^a_{hm} + \delta^b W^b_{hm}\right] + \varepsilon_{hm}
\end{aligned} \tag{3}
$$

The new binary depression indicator variable takes on the value of one for baseline CES-D scores greater than 12. When an individual who is baseline depressed receives treatment in wave w, their treatment effect from the program will be $\beta^w + \delta^w + \delta^c$ $(w = a, b)$.

Table 2 summarizes the results from estimating equation 3, displaying the expected treatment effects for both the depressed and non-depressed subpopulations. The complete regression results can be found in Appendix Table A4. Treated women with a baseline CES-D score greater than 12 experienced smaller impacts from the program on their business

Table 2: Heterogeneous Impacts by Baseline Depression, ITT Estimates

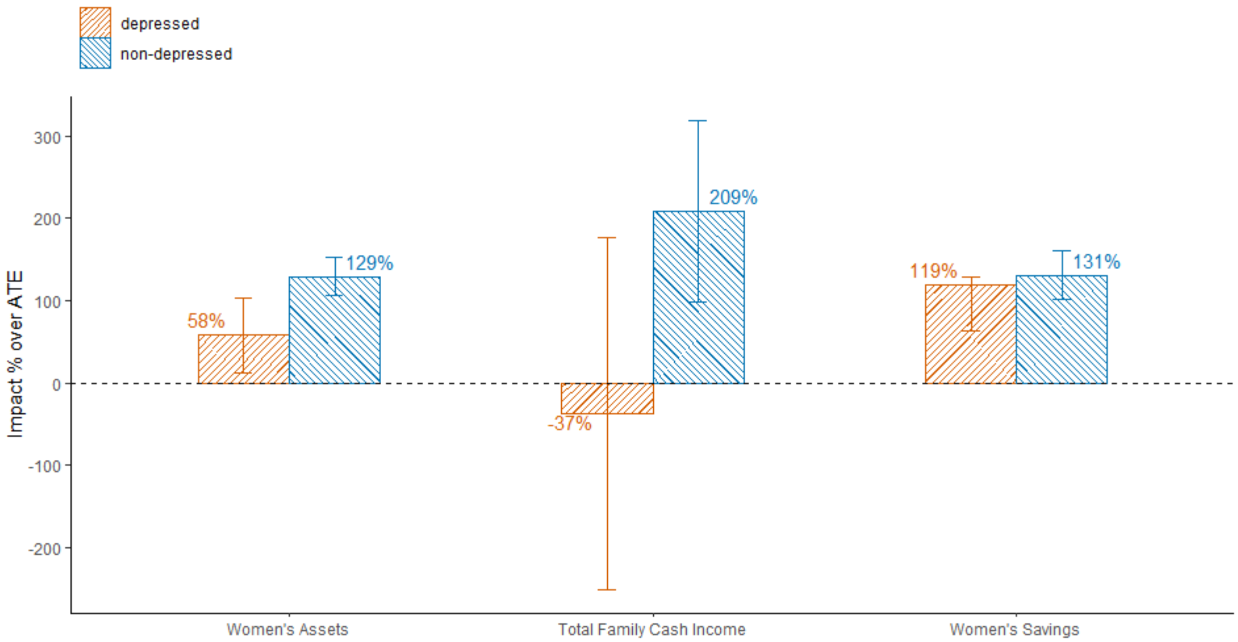
	Treatment Waves 1-2		**Treatment Waves 3-4**	
	Not Depressed	*Depressed*	*Not Depressed*	*Depressed*
Women's Business Assets ($PPP)	209***	93	137***	55
	(23.1)	(45)	(26)	(54)
Household Income ($PPP)	121	-21	2.9	4.1
	(39)	(76)	(43)	(91.2)
Women's Savings ($PPP)	56***	51	25***	17
	(7.6)	(15)	(8.5)	(17.8)
Observations		1385		

Notes: Regressions include baseline levels of the dependent variable. Standard errors for the average treatment effects are clustered at the community level. $***p < 0.01, ** p < 0.05, * p < 0.1$

assets and earnings compared to their non-depressed counterparts at baseline, resulting in a substantial divergence in outcomes between these two groups. Table A4 demonstrates that, consistent with the direct and spillover effects discussed in the previous section, participants in the REAP program experienced positive and statistically significant treatment effects. The interaction term is negative and statistically significant for Waves 1 and 2 concerning total business assets. The negative coefficient of the interaction term indicates that the group with baseline depression experienced a smaller impact on business assets compared to the group without baseline depression. This divergence suggests that the non-depressed group is accumulating more assets than the depressed group. The estimated impact for the baseline non-depressed group is $PPP 209 (se = 23.1), while for the baseline depressed group, it is $PPP 93. To facilitate a direct comparison between the impacts on the baseline depressed and non-depressed treated women, we normalized the treatment impacts on the two groups as a percentage of the intent-to-treat (ITT) impacts estimated in column 1 of Table 1. Figure 6 illustrates the comparison between the baseline depressed and baseline non-depressed groups for the three economic outcomes.

For the baseline non-depressed group, the average impact on women's assets is 129% of the average treatment effect estimated. However, for the baseline depressed women, the increase in women's assets is less than half (45%) of that. Furthermore, while the treatment impacts on family cash income for the baseline depressed women are not statistically significant, the increase in income for the baseline non-depressed women is 209% of the average treatment

Figure 6: Treatment effect between baseline depressed and baseline non-depressed groups



effects. As observed, except for savings, the treatment impacts on the non-depressed group are substantially larger. The p-values from the Wald tests assessing the equivalence of treatment effects on the depressed and non-depressed groups are 0.075 for total family cash income and 0.015 for women's assets.

It is possible that this divergence could be a result of the second grant that successful businesses received, and whether a business was successful was driven by the level of depression of the women running it. Thus, it could be that these asset accumulation patterns are reflecting the second grant. If it is true that the non-depressed were those who had successful businesses and thus received the second grant, this suggests that the depressed had a harder time benefiting from the first grant and the other components of the program, which supports the theory that the depressed are in a position that makes it harder for them to benefit from a program such as REAP.

The results also show that, though smaller compared to the non-depressed-at-baseline group, the impact of the program on business assets is nonetheless positive for the depressed-at-baseline group. The next point of interest is how an increase in assets will impact the earnings of the depressed-at-baseline group compared to the non-depressed-at-baseline group.

More specifically, in relation to the theory discussed in 2.1, the depressed-at-baseline group should be less likely to translate an increase in their assets into an increase in earnings due to the added vulnerability to poverty depression creates. In agreement with the direct and spillover impact on earnings, only Waves 1 and 2 women have experienced a statistically significant treatment effect on earnings. The coefficients on the interaction terms are not statistically significant, however they are negative and either fully or nearly cancel out the positive wave specific impact of the program on earnings for Waves 1 and 2 women. This again suggests that the depressed-at-baseline group is falling behind the non-depressed-at-baseline group in terms of earnings.

The absence of divergence in savings between these two groups may be attributed to several factors. One possibility is that non-depressed women were able to reinvest their savings into their businesses, whereas depressed women faced challenges in accumulating substantial savings. Another consideration is that savings accumulated by treated women within the savings groups underwent rigorous monitoring. Each individual group member was required to adhere to the established rules of the savings groups to access their savings. This mechanism could have functioned as a protective measure for women experiencing depression, particularly when uncertainty surrounded the specific utilization of the funds. However, it could also introduce additional inflexibility in situations requiring prompt access to funds for emergent needs. Section 6 will delve further into the implications of these findings for the design, targeting, and implementation of graduation programs.

5 Spillovers

There are a number of mechanisms by which an asset building graduation program could generate spillovers and influence others. In the first instance, there could be relatively straightforward pecuniary spillovers in which increases in the number of program beneficiaries influence the returns other individuals receive from their own economic activities. These pecuniary spillovers would be negative if more beneficiaries congests the market and lowers the prices that any individual can get for producing a good service. In our study area, many graduation program beneficiaries establish local shops or kiosks that service their local community with limited demand. Pecuniary spillovers could also be positive if they create an agglomeration economies. For example, other beneficiaries establish livestock fattening and trading businesses that service a larger market. In these cases, a growth in local businesses may make it easier to bulk purchase inputs or transport services at favorable prices.

In addition, spillovers could take place through psychosocial channels. Like other graduation programs, the BOMA REAP program intends to build both tangible physical assets as well as intangible psychosocial assets. A key difference between these two types of assets is that the former are rival goods whereas intangible assets are not, meaning that beneficiaries' psychological assets, such as self-confidence and aspirational preferences can spillover and be shared without reducing beneficiaries' stock of self-confidence or aspirations. While our reduced form identification strategy does not allow us to pin down the precise source of spillovers, we will provide evidence that at least some of the estimated reduced from spillover effects take place through psychosocial channels.

5.1 Saturation Design & Measurement

To study the impacts of spillovers from REAP participants to other REAP participants or to non-participants, the distribution of businesses started across the four treatment waves was randomly varied between communities. This was done by randomly allocating the communities across four different implementation schemes. Table 3 provides the intended distribution of communities across the four REAP distribution schemes. For example, in Scheme A communities, 60% of the REAP participants were enrolled in the first wave (March 2018), which led to relatively high saturation rates of REAP treated women (i.e., receiving the REAP training and mentoring) and REAP businesses by the 2020 data collection period. For communities in scheme B, most REAP treatments would not start until well after the 2020 survey so that there was a relatively low saturation of REAP participants and businesses in those communities during the 2020 survey. By the end of the fifth wave, all communities were to have similar saturation levels, but there was considerable variation at the time of the survey in 2020.[16]

[16] The 5th wave of REAP groups was launched just after the 2020 followup survey. By the end of wave 5, approximately 20% of all REAP-eligible women in each community had been offered the chance to participate

Table 3: Distribution of communities across saturation groups

Scheme	% of Communities	% of businesses started in each wave				
		Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
A	20%	60%	10%	10%	10%	10%
B	20%	10%	10%	10%	10%	60%
C	50%	10%	30%	20%	30%	10%
D	10%	10%	10%	60%	10%	10%

By the time of midline data collection, BOMA women started the programs in different waves had been enrolled for different durations. (*i.e.* wave 1 women would have been in the program for 24 months, wave 2 for 18 months, wave 3 for 12 months and wave 4 for 6 months). To capture the treatment duration in our saturation measures, we use a treatment-duration-weighted saturation measure (S), which we define as the probability that a random social interaction with a REAP-eligible women in the community over the 24 months between baseline and midline would have been with a woman in the REAP Program:

$$S_m = \frac{\sum_{w=1}^{4} n_m^w \times 3 \times d^w}{24 \times P_m},$$

where n_m^w is the number of businesses assigned to community m in wave w, so the number of treated women is three times that number as each business is comprised of 3 women, as described above. The d^w terms are the duration weights, which equals 24, 18, 12 and 6 for businesses established in waves 1-4, respectively. The numerator is the number of REAP-eligible women (P_m) with whom interactions are possible, weighted by the full 24-month period. Note that the weighted saturation measure S_m would be 100% if all eligible women in the community were treated in wave 1, and 0 if no women had been treated through wave 4. The mean value of this weighted saturation measure is 0.19, with a standard deviation of 0.08 and minimum and maximum values of 0 and 60%.

in REAP.

5.2 Econometric Analysis of Spillovers

To measure the intent-to-treat (ITT) impact of assignment to the REAP program and the spillover effects on both the treated and non-treated women, we estimate the following modified version of equation 2:

$$
\begin{aligned}
y_{hm} =& \alpha_0 + \alpha_1 y^0_{hm} + \beta^a W^a_{hm} + \beta^b W^b_{hm} + \\
& S_m \times \left[\delta^c I^c_{hm} + \delta^a W^a_{hm} + \delta^b W^b_{hm}\right] + + \varepsilon_{hm}
\end{aligned} \tag{4}
$$

where S_m is the community level saturation measure. I^c_{hm} is an indicator variable for an eligible woman in the control group at midline. The estimated direct impacts for a woman who enrolled in waves 1 and 2 in community m thus can be represented as: $\beta^a + \delta^a \times S_m$. The estimated spillover effects for a within-cluster non-treated woman in community m are $\delta^c \times S_m$. Following Baird et al. (2018), we define the total causal effect of being assigned to treatment wave w in community m as $\beta^w + \delta^w \times S_m + \delta^c \times (1 - S_m)$.

Table 4 presents the primary treatment and spillover results for the three key economic outcome variables that the program aimed to improve. The full regression results can be found in Appendix Table A5. Results for zero saturation are calculated by setting the saturation variable, S_m, to zero. These results thus depend on the functional form to interpolate back to the zero saturation point. To ease discussion, we focus primarily on estimated impacts for women assigned to receive treatment in waves 1 and 2. Beginning first with business assets, at the median duration-weighted saturation level of 0.19, this implies a large and statistically significant increase for those women treated in waves 1 and 2 ($PPP 240.34, se=$PPP 28.8, with a control group mean of $PPP 72.2). In addition, we see that spillovers are estimated to lead non-treated, within-cluster control households to accumulate business assets. The estimated coefficient ($PPP 270) is statistically significant and at the median duration-weighted saturation level, this implies an impact of $PPP 51.3 on the accumulation of those non-treated households. In sum, spillover impacts on the non-treated control households are about 20% of the impact on the treated.

We also find a statistically significant increase in the reported annual cash earnings

Table 4: Impacts and Spillovers at Different Saturation Levels

	Control	**Treatment Waves 1-2**		**Treatment Waves 3-4**	
	Mean Saturation	*Zero Saturation*	*Mean Saturation*	*Zero Saturation*	*Mean Saturation*
Women's Business Assets ($PPP)	51*	258***	240***	195***	171***
	(26.6)	(57.2)	(28.8)	(58)	(24)
Household Income ($PPP)	14.0	293***	137.1**	35	17
	(50.5)	(103.1)	(62.6)	(120)	(59.8)
Women's Savings ($PPP)	10.8	73**	67.2***	55**	34.5***
	(8.3)	(30)	(11.7)	(27)	(8.5)
Observations					

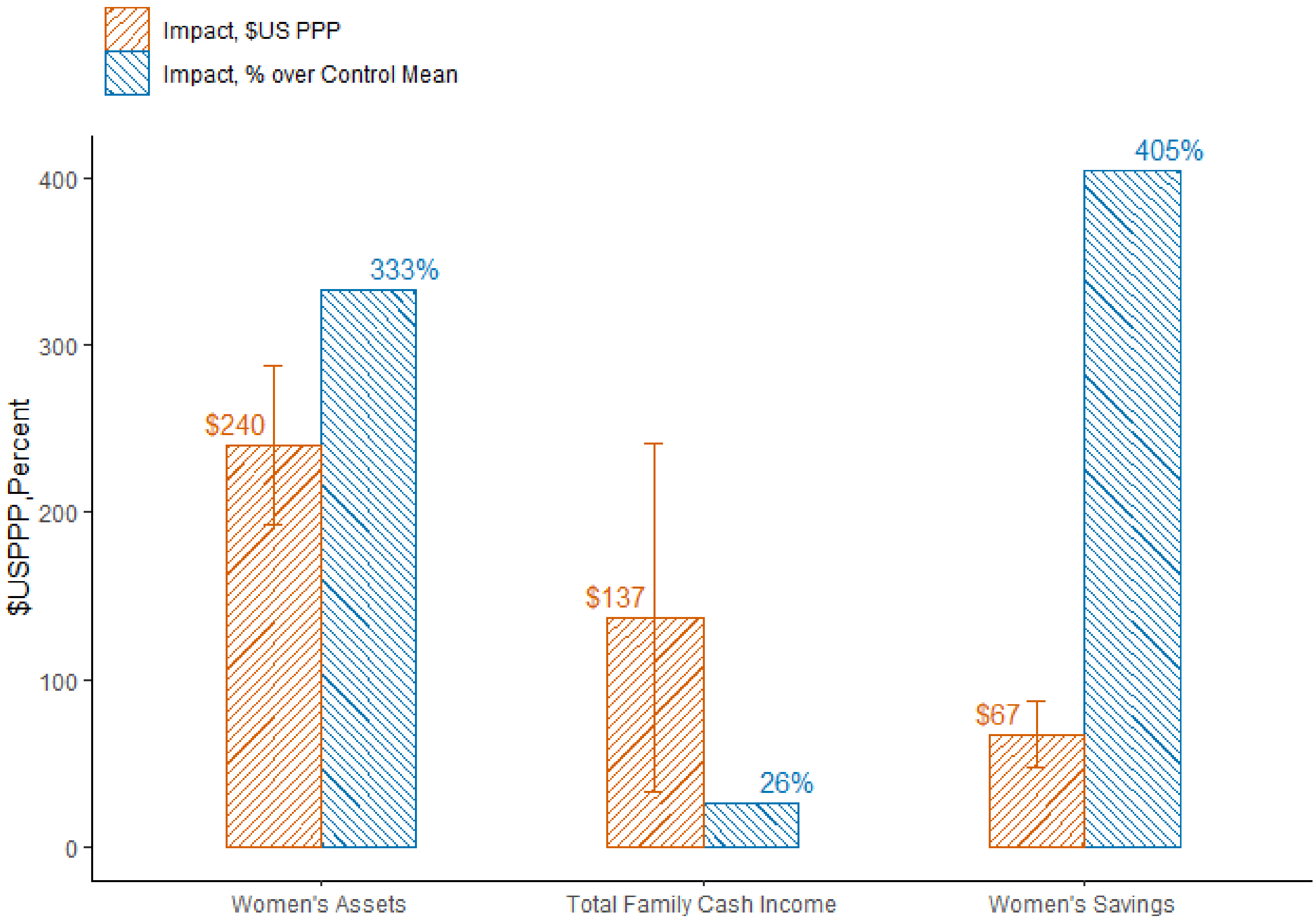
Note: Estimates are in the Appendix Table A5.

($PPP 137.1, se=$PPP 62.6, with a control group mean of $PPP 535) as well as in the amount deposited into savings ($PPP 67.16, se=$PPP 11.68, with a control group mean of $PPP 16.6). However, we observe a significant negative spillover impacts on treated households. These impacts suggest that increased competition from other REAP businesses reduces earnings. At the median duration-weighted saturation level, these estimates imply a $PPP 155.42 drop in earnings. This drop is of a magnitude that could overturn the entire amount of increase in the earnings. Since we cannot reject the hypothesis that there are no spillover effects onto the treated women, this pattern suggests that the potential impacts of REAP are reduced as more community members are treated. On the other hand, we do not observe statistically significant spillover impacts on non-treated control women for earnings.

It thus seems that while greater exposure to REAP neighbor role models spurred asset accumulation, greater competition from them led to a pattern of declining income benefits from those assets. This pattern generally holds up for both directly treated households and for non-treated households that were subject to spillover effects. The pattern for savings broadly parallels that for earnings (as would be expected if increase earnings are the source of increased savings). Impacts on savings are negative for treated and non-treated households, and in this case are uniformly statistically insignificant.

Figure 7 shows the impacts on the women who received treatment in waves 1 and 2 accounting for the spillover effects on them graphically. Note that the average value of the asset transfer of the program was $PPP 245.43 (US$100) per participant. The measured impact on the treated women including the spillover effects roughly equals to the amount

Figure 7: Causal impacts on key economic outcomes



that was transferred to them. This reflects no further accumulation of assets yet at the median saturation level villages and suggests that those treated women have managed to preserve their capital stock while generating income and savings.

To summarize, we see that increased saturation spurs asset accumulation, but lowers the income impact of those assets. One way to evaluation this tradeoff is to compare the present value of the multi-year total causal effect to the present value of the full program cost. We project out the impacts of the BOMA graduation program out over a ten year horizon.[17] Full details on this approach are given in Appendix Appendix F. While the analysis rests on a number of assumptions (namely that the benefits estimated after 24 months persist for

[17]The ten year horizon follows the analysis in Sedlmayr et al. (2020) of a graduation-like program in Uganda. While that study ignores spillovers and considers only direct impact on the treated, it does suggest a defensible way of projecting the benefits of asset-building forward into the future.

another 6 years before they begin to dissipate), it does allow us to gauge the impact of the program as a function of the saturation rate. We also use the Table 1 average treatment effects to calculate the benefit-cost ratio we would have obtained had we ignored spillovers. From the perspective of the implementing program, these benefit-cost numbers represent their return on investment in the poor population in the pastoralist regions.

The results of this analysis are as follows. Ignoring spillovers, we obtain a benefit-cost ratio of 1.7. In words, every dollar invested in the REAP program generated $1.7 in benefits for REAP eligible women. That same measures rises to 2.2 when we evaluate impacts at the average saturation level of 19%. Finally, if we calculate the total causal effect at a lower saturation ratio of only 15%, then the benefit-cost ratio rises to just over 3. While program design is more complex than maximizing the benefit-cost ratio, these data indicate that a lower saturation level may be called for in the program being evaluated.

5.3 Do Adaptive Preferences Explain Spillovers?

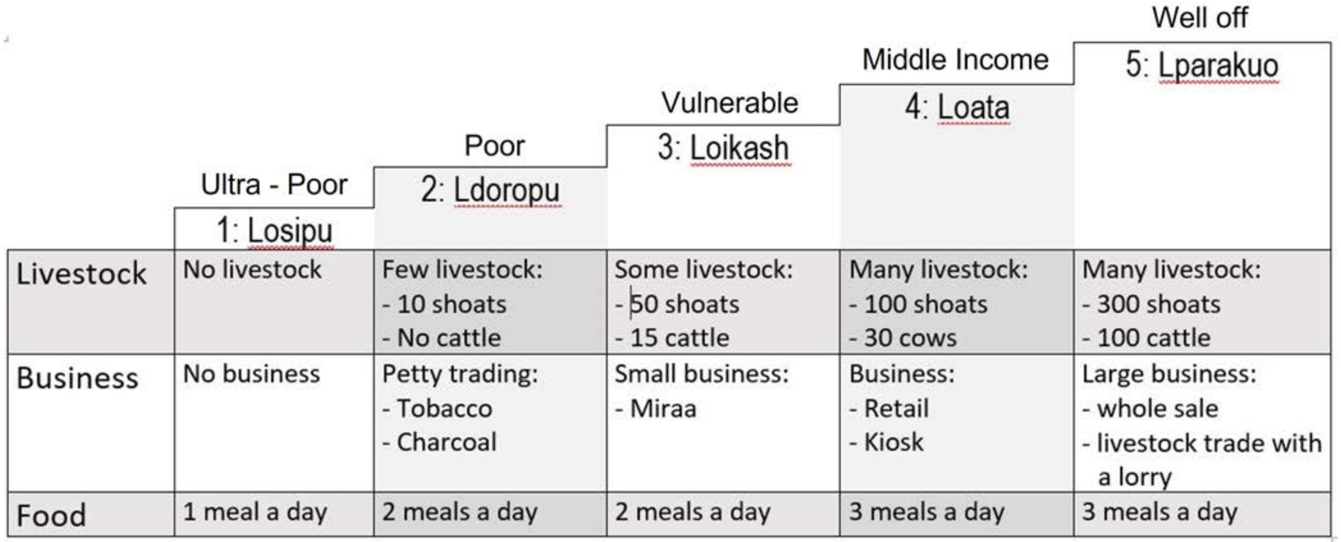
While the negative impacts of more businesses on cash earnings is not necessarily surprising given that most BOMA program beneficiaries live in small, remote communities where it is easy to imagine saturating the market,[18] the more intriguing result is the finding that women who were not in the REAP program began to changed their behavior and accumulate tangible assets. While this behavioral change could be explained by simple exogenous social effects (*e.g.*, asset accumulation became more profitable when neighbors began to build up their own businesses and income)[19], we here explore Section 2.4's suggestion that treated women showed their untreated peers that women could start businesses and obtain higher living standards for themselves and their families. In the notation of Section 2.4, this demonstration effect shifted out the perceived income ceiling, $\tilde{c}$, inducing an increase in the perceived marginal utility of economic advancement, as suggested by the theory of adaptive preferences.

While measurement of the subjective marginal utility of income is non-trivial, we built our approach on a five-step ladder of life that portrayed different levels of living standards of

[18]Indeed, prior to this study, the REAP program purposefully limited the number of women treated in any community to no more than one-third of the eligible population.

[19]See discussion in Manski (1993) on social effects

Figure 8: Ladder of Life for Pastoralist Population



	1: Losipu	2: Ldoropu	3: Loikash	4: Loata	5: Lparakuo
Livestock	No livestock	Few livestock: - 10 shoats - No cattle	Some livestock: - 50 shoats - 15 cattle	Many livestock: - 100 shoats - 30 cows	Many livestock: - 300 shoats - 100 cattle
Business	No business	Petty trading: - Tobacco - Charcoal	Small business: - Miraa	Business: - Retail - Kiosk	Large business: - whole sale - livestock trade with a lorry
Food	1 meal a day	2 meals a day	2 meals a day	3 meals a day	3 meals a day

local communities.[20] This ladder of life approach is adapted from the Cantril ladder, which asks respondents to evaluate their current life as a whole using the mental image of a ladder. The Gallup World Poll, which remains the principal source of data in the World Happiness Report (Helliwell et al. (2022)), uses the responses to this question as life evaluations of the measurement of subjective well-being. We characterized each step of the ladder over three dimensions: livestock, business and food based on community understandings of different standards of living for who is destitute, poor, vulnerable, middle income and well-off. Respondents were then asked to place themselves on the ladder described using Figure 8 (most on steps 1 & 2) and told us how important (on a scale from 1 to 5) it was to work hard to advance to each one of the higher steps. We then standardize the importance index to the next steps to be our measure on respondents' beliefs on the importance of moving up the social ladder. The underlying assumption is that the importance assigned by any respondent to moving up the ladder is an analogue measure of marginal utility.

Using the same empirical approach outlined in the spillover equation 4in Section 5.2, we examine the direct and spillover effects of the REAP program on individuals' valuations of upward mobility in life. As depicted in Table 5, our analysis reveals that, in comparison to a pure control group, the treatment exerts a substantial and statistically significant influence

[20]The REAP program uses this same ladder approach in its participatory poverty assessment that is used to determine graduation program elibigibility.

(0.48 standard deviations, se = 0.21) on the aspiration to progress to step 3, accompanied by a slightly smaller impact (0.37 standard deviations, se = 0.19) on advancing to step 4. Furthermore, we note a similar albeit diminished impact on control women (0.84 standard deviation increase, se = 0.48 for step 4). We present the Ordinary Least Squares (OLS) regression results, as the interpretation of the coefficients remains relatively straightforward. Additionally, the results of the ordered probit regression are provided in the appendix, under Table A7, utilizing the importance scale (ranging from 1 to 5, with 5 indicating high importance) as the dependent variable, yielding qualitatively consistent outcomes. The overarching goal of graduation programs like REAP is to enhance the resilience of program participants by alleviating constraints. Furthermore, our findings indicate that the program not only influences participants' desires, but also lead to an increase in their neighbors' assets, as discussed above. This increase is consistent with a behavioral change induced by the endogenous change in preferences (as analyzed theoretically in section 2.4). We cannot rule out other explanations, such as gift-giving from treated to non-treated women.

triggers a social spillover or demonstration effect among their neighbors. While we have yet to ascertain the exact proportion of the estimated impacts attributable to social multipliers, a discernible pattern is emerging between these shifts in preferences and the cascading effects on capital accumulation.

6 Conclusion

The arid and semi-arid pastoralist regions of the Horn of Africa constitute one of the most challenging environments for graduation programs attempt to reduce poverty by build productive assets for women. We indeed confirm that the variant of the graduation program model developed by the BOMA Project NGO indeed works and exhibits a benefit-cost ration in excess of 2.

However, the goal of this evaluation steps beyond simply stress-testing the graduation model in difficult environment. In particular, the evaluation was designed to allow exploration of the psychosocial channels through which graduation programs operate. Drawing on equal measures of the economics of poverty traps (Ikegami et al., 2019) and the economics

Table 5: Regression results on Ladder of Life

VARIABLES	Importance to get to ladder 3	Importance to get to ladder 4	Importance to get to ladder 5
ITT Wave 1 & 2	0.48**	0.37*	0.30*
	(0.21)	(0.19)	(0.18)
ITT Wave 3 & 4	0.14	-0.13	-0.28
	(0.25)	(0.25)	(0.22)
ITT w1-w2 * Saturation	-1.08	-0.96	-0.68
	(0.7)	(0.75)	(0.67)
ITT w3-w4 * Saturation	1.08	2.12**	2.43**
	(0.99)	(1.07)	(1.08)
Control * Saturation	0.62	0.84*	0.57
	(0.62)	(0.48)	(0.52)
Baseline level of outcome variables	-0.017	0.02	0.012
	(0.029)	(0.034)	(0.024)
Constant	-0.13	-0.15	-0.099
	(0.11)	(0.092)	(0.097)
Observations	830	1,353	1,382
R-squared	0.013	0.009	0.007

Note: Standard errors clustered at the community level. Importance scale (1-5 very important) standardized at control mean.*** p<0.01, ** p<0.05, * p<0.1.

of depression (de Quidt and Haushofer, 2016), we show that consistent with theory the impressive average treatment effects are primarily driven by beneficiaries who exhibited strong baseline mental health. The 20% of the beneficiaries with severe depressive symptoms in fact benefit not at all from the program, drawing down on the assets transferred to them and experiencing no income gain. While we cannot claim that this finding explains the often observed heterogeneity in impacts in which 25-30% of beneficiaries experience no benefits (e.g., see Bandiera et al. 2017 and Gobin et al. 2017), the consistency of the theory and empirical evidence is suggestive.

The second psychosocial mechanism we explore is an endogenous preference mechanism (adaptive, or sour grapes preferences modeled on (Elster, 1983)) that is shown theoretically to generate substantial behavioral change in untreated population. Using the empirical study's randomized saturation design, we are able to document spillovers from the treated to untreated that result in increased business asset accumulation by the latter. We further find evidence that increased exposure to treated women increases the subjective value that women assign to economic advancement. These spillover effects are economically substantial and they increase the estimated benefit-cost ratio of the program from 1.7 to 2.2. There is also evidence that the existing program modestly over-saturates the remote communities in the study area as a lower saturation rate would increase the benefit-cost ratio to just over 3.

A final observation is that these reported benefit-cost numbers include the drag of the women who lacked the psychological assets to benefit from the program. One implication is that the program could generated more benefits per-dollar expended if additional potential beneficiaries were screened for the kind of severe depressive symptoms that seem to limit program impact. Another, and perhaps more palatable implication, would be to design a two-track program in which women without severe depressive symptoms would move forward with the program as currently designed while those showing depressive symptoms would take a different track. Recent work on cognitive behavioral therapy reported in Barker et al. (2022) suggests that low cost interventions may be able to resolve the mental health issues that afflict other women. Resolving those health issues first, in a second track graduation program, may lead to better average outcomes, while still making those benefits available to all women who need them based on an economics means test.

In closing, as is clear from the theory of poverty traps analyzed in Section 2, another source of impact heterogeneity stems from different exposure to shocks. In forthcoming analysis, we use a second follow-up survey to study the impact of an index insurance contract to ward off the ill effects of shocks on women's ability to preserve asset built by the BOMA program.

References

Appadurai, Arjun, "The Capacity to Aspire: Culture and the Terms of Recognition," in "Rao, V. and Walton, M., (eds.) Culture and Public Action," Stanford University Press, Palo Alto, California,, 2004, pp. pp59–84.

Baird, Sarah, J. Aislinn Bohren, Craig McIntosh, and Berk Özler, "Optimal Design of Experiments in the Presence of Interference," *The Review of Economics and Statistics*, December 2018, *100* (5), 844–860.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman, "Labor markets and poverty in village economies," *The Quarterly Journal of Economics*, 2017, *132* (2), 811–870. Publisher: Oxford University Press.

Barker, Nathan, Gharad Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher Udry, "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress," *American Economic Review: Insights*, December 2022, *4* (4), 527–545.

Baron, Emily Claire, Thandi Davies, and Crick Lund, "Validation of the 10-item Centre for Epidemiological Studies Depression Scale (CES-D-10) in Zulu, Xhosa and Afrikaans populations in South Africa," *BMC Psychiatry*, 2017, *17* (1), 6. Publisher: BMC Psychiatry.

Beck, Aaron T, *Depression: Causes and Treatment*, 1 ed., Philadelphia: University of Pennsylvania Press, 1967.

Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand, Catherine C. Thomas, Christopher Udry, Julia Vaillant, and Kelsey A. Wright, "Tackling psychosocial and capital constraints to alleviate poverty," *Nature*, May 2022, *605* (7909), 291–297. Number: 7909 Publisher: Nature Publishing Group.

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin, "Macro-Perspective on Asset Grants Programs: Occupational and Wealth Mobility," *American Economic Review*, May 2014, *104* (5), 159–164.

Christian, Cornelius, Lukas Hensel, and Christopher Roth, "Income Shocks and Suicides: Causal Evidence From Indonesia," *The Review of Economics and Statistics*, 2019, *101* (5), 905–920.

de Quidt, Jonathan and Johannes Haushofer, "Depression for Economists," December 2016.

Elster, Jon, *Sour grapes : studies in the subversion of rationality*, Cambridge [Cambridgeshire] ;: Cambridge University Press, 1983. Publication Title: Sour grapes : studies in the subversion of rationality.

Genicot, Garance and Debraj Ray, "Aspirations and Inequality," *Econometrica*, 2017, *85* (2), 489–519. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA13865.

Gobin, Vilas J., Paulo Santos, and Russell Toth, "No Longer Trapped? Promoting Entrepreneurship Through Cash Transfers to Ultra-Poor Women in Northern Kenya," *American Journal of Agricultural Economics*, October 2017, *99* (5), 1362–1383.

Hashemi, Syed M, "Reaching the Poorest: Lessons from the Graduation Model," 2011.

Helliwell, John F, Shun Wang, Haifang Huang, and Max Norton, "Happiness, Benevolence, and Trust During COVID-19 and Beyond," 2022, p. 40.

Hulme, David and Karen Moore, "Social Protection for the Poor and Poorest," in Armando Barrientos and David Hulme, eds., *Armando Barrientos and David Hulme, eds.*, Palgrave Macmillan UK, 2008. Section: Assisting the Poorest in Bangladesh: Learning from BRAC's 'Targeting the Ultra-poor' Programme.

Ikegami, Munenobu, Michael R. Carter, Christopher B. Barrett, and Sarah Janzen, "6. Poverty Traps and the Social Protection Paradox," in "6. Poverty Traps and the Social Protection Paradox," University of Chicago Press, January 2019, pp. 223–262.

Kilburn, Kelly, Leah Prencipe, Lisa Hjelm, Amber Peterman, Sudhanshu Handa, and Tia Palermo, "Examination of performance of the Center for Epidemiologic Studies Depression Scale Short Form 10 among African youth in poor, rural households," *BMC Psychiatry*, June 2018, *18*, 201.

Laajaj, Rachid, "Endogenous time horizon and behavioral poverty trap: Theory and evidence from Mozambique," *Journal of Development Economics*, July 2017, *127*, 187–208.

Lybbert, Travis J. and Bruce Wydick, "Poverty, Aspirations, and the Economics of Hope," *Economic Development and Cultural Change*, 2018.

Manski, Charles F., "Identification of Social Endogenous Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

Montgomery, James, "Revisting Tally's Corner: Mainstream Norms, Cognitive Dissonance and Underclass Behavior," *Rationality and Society*, 1994, *6* (4), 462 – 488.

Radloff, L S, "The CES-D Scale: A Self Report Depression Scale for Research in the General," *Applied Psychological Measurement*, 1977, *1*, 385–401. ISBN: 0146-6216.

Rotter, Julian B., "Generalized expectancies for internal versus external control of reinforcement," *Psychological Monographs: General and Applied*, 1966, *80* (1), 1–28. Place: US Publisher: American Psychological Association.

Sedlmayr, Richard, Anuj Shah, and Munshi Sulaiman, "Cash-plus: Poverty impacts of alternative transfer-based approaches," *Journal of Development Economics*, May 2020, *144*, 102418.

World Heath Organization, "Depression and Other Common Mental Disorders: Global Health Estimates," Technical Report, World Health Organization, Geneva 2017.

Appendices For Online Publication

Appendix A Numerical Parameterization of Occupational Choice Model

Table A1: Functional Forms and Parameters used in Numerical Simulations

Production Technology and Parameters
$F^w_{jt} = w_0 + k_{jt}^{\gamma_L}$ $F^e_{jt} = (w_0 - A) + \alpha_j k_{jt}^{\gamma_H}$ $\gamma_L = 0$ $\gamma_H = 0.56$ $A = 3.95$ $w_0 = 3.95$
Utility Function and Parameters
Adaptive preferences utility function: $u(c_{it}) = \begin{cases} u^{\ell}(c_{it})\ if\ c_{it} < \tilde{c}(c_{g(i)}) \\ u^h(c_{it})\ otherwise \end{cases}$ Conventional prefeneces utility function: $u^l(c_t) = \frac{c_t^{1-\rho_l}-1}{1-\rho_l}$ $\beta = 0.95$ $\rho_l = 0.75$ $\rho_h = 2.5$
Distribution of Shocks
The probability of θ_{jt} is assumed to be: density of $\theta_{jt} = \begin{cases} 0.3 & \theta_{jt} = 0.11 \\ 0.18 & \theta_{jt} = 0.021 \\ 0.13 & \theta_{jt} = 0.031 \\ 0.11 & \theta_{jt} = 0.041 \\ 0.10 & \theta_{jt} = 0.051 \\ 0.02 & \theta_{jt} = 0.061 \\ 0.01 & \theta_{jt} = \{0.071, 0.081, ..., 0.191\} \end{cases}$

Appendix B Balance and Attrition Tables

Table A2: Baseline Balance

Panel A: T-test comparing means of baseline characteristics by endline treatment status				
	Control mean [standard error]	Treatment mean [standard error]	p-value from t-test	Normalized difference
Baseline Reported Annual Cash Earnings (KES)	33313.454	36319.322	0.071*	-0.097
	[1100.568]	[1258.628]		
Baseline Total Business Assets (KES)	1880.193	2242.171	0.5	-0.036
	[290.374]	[472.487]		
Baseline Amount Deposited Into Savings (KES)	466.494	665.586	0.212	-0.067
	[90.057]	[137.083]		
Tropical Livestock Units BL	3.108	3.057	0.796	0.014
	[0.128]	[0.156]		
Baseline Household Dietary Diversity Score	3.1	3.019	0.166	0.075
	[0.040]	[0.042]		
Ladder of Life Step	2.089	2.111	0.446	-0.041
	[0.019]	[0.021]		
Locus of control scores (higher - more external)	33.883	33.432	0.214	0.067
	[0.248]	[0.263]		
Baseline Internal Locus of Control Score	17.278	17.154	0.291	0.057
	[0.080]	[0.085]		
Baseline CES-D	8.397	8.429	0.901	-0.007
	[0.171]	[0.200]		
Baseline importance to move to ladder 3	4.579	4.585	0.841	-0.012
	[0.021]	[0.023]		
Baseline importance to move to ladder 4	4.643	4.65	0.82	-0.012
	[0.019]	[0.019]		
Baseline importance to move to ladder 5	4.731	4.69	0.133	0.081
	[0.018]	[0.021]		
Other people sharing business knowledge	-0.016	-0.353	0.241	0.063
	[0.132]	[0.272]		
Age of HH head	43.821	43.342	0.618	0.028
	[0.643]	[0.716]		
Age of respondent	36.101	35.159	0.286	0.058
	[0.605]	[0.639]		
% respondent with husbands	0.612	0.583	0.273	0.059
	[0.018]	[0.020]		
HH head yrs of schooling	1.195	1.177	0.916	0.006
	[0.112]	[0.125]		
HH size	5.437	5.331	0.345	0.051
	[0.076]	[0.081]		
	# of obs control	# of obs treatment		
Panel B: Regression of Treatment on all outcomes				
F-test from regression of treatment on all outcome variables listed above	1.49			
p-value	0.11			

Table A3: Attrition: Dependent Variable: Completed Survey, OLS

	Midline	Endline
Panel A		
Treatment Status	-0.0071	-0.00013
	(0.013)	(0.017)
Observations	1,874	1,874
R-squared	0.052	0.063
Outcome mean	0.92	0.86
Panel B		
Treatment Status	-0.0059	0.0012
	(0.013)	(0.017)
Household Reported Cash Earnings(KES)	8.3e-08	2.3e-07
	(1.1e-07)	(1.5e-07)
Total Business Assets(KES)	7.1e-08	-2.1e-07
	(3.1e-07)	(4.1e-07)
Savings(KES)	4.6e-07	-1.6e-06
	(1.4e-06)	(1.8e-06)
Household Dietary Diversity Score	-0.00021	0.0026
	(0.0063)	(0.0083)
Ladder of Life Step	-0.0035	-0.010
	(0.025)	(0.033)
Locus of Control	-0.00018	-0.00022
	(0.0011)	(0.0014)
Internal Locus of Control	-0.00073	-0.00046
	(0.0033)	(0.0043)
Shared with Business Knowledge	-0.050**	-0.033
	(0.0014)	(0.0018)
CES-D Score	-0.0042***	-0.0039**
	(0.0014)	(0.0018)
TLU (Tropical Livestock Units)	-0.00072	-0.0013
	(0.0018)	(0.0024)
Importance to move to ladder 3	0.0089	0.0062
	(0.017)	(0.023)
Importance to move to ladder 4	-0.0036	-0.00042
	(0.019)	(0.024)
Importance to move to ladder 5	-0.0049	0.011
	(0.018)	(0.023)
Observations	1,874	1,874
R-squared	0.062	0.070
Outcome mean	0.92	0.86
Panel C		
Treatment Status	0.36	0.50
	(0.25)	(0.32)
Baseline characteristics	YES	YES
Baseline characteristics interacted with Treatment	YES	YES
Observations	1,874	1,874
R-squared	0.069	0.080
Outcome mean	0.92	0.86
P-value from test that Treatment and all other variables above interacted with Treatment are jointly 0	0.66	0.17

Appendix C Imputing the missing data on BOMA total business assets

During the midline data analysis, we encountered a discrepancy between the reported total business assets from the survey responses of women assigned to the treatment group and the BOMA administrative data from February 2020. To investigate this further, in March 2021, we conducted a follow-up round of discussions with BOMA mentors to understand the reasons behind these discrepancies.

In summary, out of the total sample of 241 women, 8 (approximately 3%) were confirmed as attrition cases, either due to non-participation at the time of the survey or never having participated in the program. Among the remaining 233 women, around 14% (33 women) experienced various disruptions during the midline data collection period, such as forest evictions, security concerns, or temporary travel/migration.

According to the mentors' feedback, there were several reasons for the reported zero business assets. First, some participants did not realize that the mentors were referring specifically to their BOMA businesses, leading to confusion in their responses. Second, a significant proportion of women (approximately 66% based on the provided Excel sheet) either did not fully understand the question or had difficulty comprehending it. Additionally, some women admitted to not being fully focused during the interview or feeling shy, which could have affected their responses. Among the women, 10% (23 women) were mentioned by mentors as being less active or not directly involved in the day-to-day operations of the business during the midline survey period. Finally, mentors suggested that a subset of women (approximately 12%, or 27 women) intentionally withheld information from the enumerators, even if they understood the questions.

Imputation of missing values

For the cases where participants reported zero business assets but had business value recorded in the BOMA administrative data, we made a change in our analysis. Specifically, we marked the zero total business assets as missing for those participants who were identified by the BOMA mentors as "Did not understand the question" only. Among the total sample of 241 cases, 135 (approximately 56%) were identified by the BOMA mentors as having non-zero business assets.

However, we did not make any changes to the reported zeros for the other categories of participants. This decision was made because we observed instances of zero business assets reported by the control group, even when they self-identified as owning businesses. Additionally, we lacked information regarding the reasons behind the control group's reported zero business assets. Therefore, we chose not to alter the reported zeros for these categories during the analysis.

Predictive mean matching

We used predictive mean matching to impute those missing values. In STATA, it is under multiple imputation method. Predictive mean matching (PMM) is a partially parametric method that matches the missing value to the observed value with the closest predicted mean (or linear prediction). It was introduced by Little (1988) based on Rubin's (1986) ideas applied to statistical file matching. PMM combines the standard linear regression and the nearest-neighbor imputation approaches. It uses the normal linear regression to obtain linear predictions. It then uses the linear prediction as a distance measure to form the set of nearest neighbors (possible donors) consisting of the complete values. Finally, it randomly draws an imputed value from this set. By drawing from the observed data, PMM preserves the distribution of the observed values in the missing part of the data, which makes it more robust than the fully parametric linear regression approach.[21]

[21] For detailed information on this step: STATA manual mi impute pmm Methods and formulas.

Appendix D Impact Regression Tables

Table A4: Full Regression Results

	Standard ITT			Baseline Depression Heterogeneity		
	Business Assets	*Family Income*	*Savings*	*Business Assets*	*Family Income*	*Savings*
Waves 1 & 2	190***	98.2***	56.3***	209***	121***	56.0***
	(22.3)	(34.2)	(8.27)	(23.1)	(39.2)	(7.64)
Waves 3 & 4	125***	4.43	25.0***	137***	2.89	25.1***
	(18.6)	(36.3)	(6.19)	(25.6)	(43.4)	(8.45)
Waves 1 & 2 × Depression				-111**	-133	1.77
				(55.9)	(94.6)	(18.4)
Waves 3 & 4 × Depression				-76.9	10.6	-1.19
				(64.4)	(109)	(21.2)
Depression				-4.74	-9.37	-6.66
				(29.7)	(50.3)	(9.79)
Baseline Level Dep. Variables	0.13*	0.11***	0.20**	0.13***	0.10***	0.20***
	(0.064)	(0.029)	(0.092)	(0.036)	(0.020)	(0.040)
Constant	66.2***	449***	14.5***	66.9***	452***	15.7***
	(10.7)	(28.0)	(2.94)	(12.3)	(26.2)	(4.03)
Observations	1,385	1,385	1,385	1,385	1,385	1,385
R^2	0.071	0.027	0.066	0.076	0.030	0.066

Notes: Baseline levels of the outcome of interest and an indicator for vulnerable group are included in both specifications for ITT and quantile regressions. Total income was measured by adding reported cash income from sales of livestock, livestock products, crops, casual labor, salaried employment, and business, duka, and petty trading for each household. Assets were calculated by summing up the cash, stock, assets, savings, and credits related to each individual business that the respondents own. Standard errors clustered at the community level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix E ATE on Psychological Variables

Table A6 presents results on psychological outcomes. While accounting for spillovers does not change the result that the REAP program had no impact on women's CES-D score. We do see a consistent pattern of positive program impacts on mental health (lower numbers mean less likely to be depressed, and lower numbers means more internal locus of control) and improvement in the life evaluation on the Waves 1 and 2 women (measured as the current step on the ladder of life). Nor did we find any statistically significant improvement on the within-cluster control women on the CES-D score. On locus of control, which measures the degree to which people believe that they have control over the events that affect their lives — internal locus of control —, as opposed to external forces — external locus of control — (Rotter (1966)). We find that the REAP program reduces the locus of control measure for the participants (which moves them towards more internal locus of control). However,

Table A5: Impact Estimates Accounting for Spillovers

VARIABLES	Total Business Assets ($PPP)	Reported Annual Cash Earnings ($PPP)	Amount Deposited Into Savings ($PPP)
ITT Wave 1 & 2	258***	293***	72.6**
	(57.2)	(103)	(30.4)
ITT Wave 3 & 4	195***	35.4	54.9**
	(58.0)	(120)	(27.5)
ITT w1-w2 * Saturation	-92.1	-818**	-28.7
	(220)	(409)	(118)
ITT w3-w4 * Saturation	-123	-96.7	-107
	(283)	(523)	(122)
Non-treated Control * Saturation	270*	73.7	57.0
	(140)	(266)	(43.6)
	[0.207]	[0.414]	[0.243]
Baseline Level of Dependent Variables ($PPP)	0.12*	0.11***	0.20**
	(0.064)	(0.029)	(0.092)
Constant	19.2	436***	4.59
	(19.3)	(54.2)	(6.57)
Observations	1,385	1,385	1,385
R-squared	0.073	0.030	0.067
Control Mean of Dependent Variable	72.2	535	16.6

Note: Standard errors clustered at the community level. Sharpened q-values for estimated coefficients on within-cluster control reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Treatment effects on secondary outcomes

VARIABLES	CES-D	Locus of Control	Life Evaluation
ITT Wave 1 & 2	-0.018	-0.34*	0.35**
	(0.18)	(0.18)	(0.15)
ITT Wave 3 & 4	-0.31	-0.42	-0.13
	(0.19)	(0.27)	(0.24)
ITT w1-w2 * Saturation	-0.32	1.32*	-0.69
	(0.73)	(0.79)	(0.57)
ITT w3-w4 * Saturation	1.27	2.59*	1.31
	(0.82)	(1.38)	(1.35)
Non-treated Control * Saturation	-0.073	0.14	0.27
	(0.47)	(0.49)	(0.48)
Baseline level of outcome variables	0.11***	-0.019	0.23***
	(0.028)	(0.027)	(0.034)
Constant	0.013	-0.025	-0.048
	(0.093)	(0.095)	(0.088)
Observations	1,384	1,378	1,385
R-squared	0.015	0.008	0.061

Note: Standard errors clustered at the community level. Outcome variables standardized at control mean. *** p<0.01, ** p<0.05, * p<0.1.

accounting for the spillover effects undercut the direct effects and we find no sign of spillovers to control on locus of control, either.

Appendix F Benefit-Cost Methodology for Total Causal Effects

One way to gauge the effectiveness of the REAP program and these individual impacts is to calculate the estimated discounted present value of all benefits (direct & indirect) and compare those benefits to the present value of the program cost. We evaluate the return on investment ratio (ROI) taking into account this admittedly complex pattern of spillovers and the fact that non-treated households receive spillover benefits without any additional

Table A7: Ordered Probit results on Ladder of Life

VARIABLES	Importance to get to ladder 3	Importance to get to ladder 4	Importance to get to ladder 5
ITT Wave 1 & 2	0.654**	0.491**	0.401**
	(2.35)	(2.18)	(1.96)
ITT Wave 3 & 4	0.178	-0.087	-0.309
	-0.52	-0.31	-1.11
ITT w1-w2 * Saturation	-1.5	-1.375	-0.874
	(1.56)	(1.55)	(1.15)
ITT w3-w4 * Saturation	1.598	2.38*	3.088**
	(1.12)	(1.90)	(2.05)
Control * Saturation	0.991	1.015*	0.875
	(1.26)	(1.71)	(1.38)
Baseline level of outcome variables	-0.051	0.051	0.035
	(0.75)	(0.63)	(0.51)
Observations	830	1,353	1,382

Note: Standard errors clustered at the community level. Importance scale (1-5 very important) estimated using ordered probit.*** p<0.01, ** p<0.05, * p<0.1.

expenditure. We calculate the ROI using the following formula:

$$(5) \quad ROI(S) = \sum_{t=1}^{10} \left[\frac{TCE_t(S)}{n \times \widetilde{S} \times c_t} \right] = \sum_{t=1}^{10} \left[\frac{S(\beta^w + S\delta^w) + (1-S)(S \times \delta^c)}{\tilde{S} \times c_t} \right]$$

where c is the cost per-treated person. Note that the TCE(S) in the numerator is, as before, the total benefits to a community of n eligible people where a fraction of S are treated. The denominator gives the investment needed to receive that impact (number of people treated, $n \times \widetilde{S}$ times the cost per-fully treated person. The new term $\tilde{S}$ is a slightly modified version of the duration weighted saturation measure that better captures the relative expense of treating individuals in waves 1, 2 and 3 compared to wave 4.[22] The expanded expression of the ROI expression shows that the ROI is just the per-capita benefit (weighted average of

[22] Costs of treating a REAP beneficiary is not quite proportional to time in the program. For example, a wave 4 woman was in REAP for only 25% of the time as a wave 1 women, but the expenditures on a wave 1 woman was about 37% of the cost of a wave 1 woman. The new term \tilde{S} replaces the time weights used to define S with these relative expenditure weights.

direct and indirect benefits) divided by the average cost of all people treated $(s \times c + (1-s) \times 0)$. Note further that if there are no spillover impacts and $\delta^{w,c} = 0$, then this expression reduces down to the more standard looking $\frac{\beta^w}{c}$.

Using the cost per household converted to $PPP 713.21, the benefit/cost ratio is 2.2:1, estimated from BOMA's point of view. In words, every dollar invested in the REAP program generated $2.2 in benefits for REAP eligible women. This increase compared to the simple measure had we estimated a naive ITT results from valuing the spillovers to non-treated eligible women, which in turn causes some undercounting of the benefits to treated women. We have not considered impacts on women in BOMA communities who are not eligible for the program.

We take total causal effect estimated for reported annual cash earnings for wave 1 and wave 2 participants and assume that those treatment effects persist constant until the end of seven years after the intervention begins, then gradually decline over the next three years. While we do not model any changes to business assets or savings through years 2 through 7, starting in year 8 we assume that the participants liquidate their savings and business assets over a period of three years as their businesses wind down. We also assume that any other differences between participants and nonparticipants because of the REAP program disappears by the 10-year mark. We assume that one third of the savings and asset stocks generated (and presumably held) enter as cash flow in year 8, one third in year 9 and the final third in year 10. We assume a discount rate of 5% per year. At the median duration-weighted saturation level of 0.19, the total discounted value of the intervention is estimated to be $PPP 350.99 by this approach. According to BOMA's estimation on cost to deliver the program, the cost per participant for Wave 1 participants (who had graduated from the program by midline) is $322.46 per household (after discounting to the present value). This number includes all program delivery, managerial and administrative costs associated with the REAP program, as reported by BOMA administrative data.