

Online Supplement for “Aspiring to a Better Future: Can a Simple Psychological Intervention Reduce Poverty?”

Kate Orkin, Robert Garlick, Mahreen Mahmud,
Richard Sedlmayr, Johannes Haushofer, Stefan Dercon

December 11, 2024

Contents

A	Additional Information about Workshop Interventions	2
A.1	Summaries of Video Content	2
A.2	Exercise Scripts	3
A.3	Administration of Workshops	7
B	Departures from Preanalysis Plan	7
B.1	Measurement	7
B.2	Econometric Specification	8
B.3	Summary Indices and Adjustments for Multiple Hypothesis Testing	9
B.4	Data Collected Directly After Intervention	9
C	Details on Cost and Benefit Calculations	10
D	Details on Measurement	12
D.1	Economic Measures	12
D.2	Psychological Measures	15
D.3	Development and Validation of Psychological Scales	20
E	Additional Results	22
E.1	Heterogeneous Treatment Effects	22
E.2	Associations Between Wealth, Investment, Expectations, and Other Psychological Characteristics.	26

A Additional Information about Workshop Interventions

A.1 Summaries of Video Content

All videos are posted at <https://mbrg.bsg.ox.ac.uk/aspirations-and-goal-setting-video-intervention>.

Aspirations: These tell stories about the lives of Judy and Josefine. Judy and her partner Oyoo are smallholder farmers with few assets, and are expecting a child. They discuss struggling to get by and what they want for the future. They decide that within five years, they want to put an iron roof on their house, for their children to complete school and for Judy to start a business. They set intermediate goals to save 100 shillings each week, grow more vegetables to sell at market, and purchase a plot within a year. They put money in a small container to save. Judy learns to sew, overcoming some obstacles to do so, and starts a successful business making clothes. They succeed in buying an iron roof. She adjusts her business plan to deal with competition from cheap imported clothing and her business prospers. Eventually, they send their child to university.

Josefine is a teacher and farmer. She tells her life story of how she came to be successful. She begins by remembering that she used to beg for money as a child and work as a casual day labourer. Her teacher describes how she dropped out of school several times. Another woman describes teaching Josefine to weave baskets to sell at market. Her husband describes how she saved money from this to go to high school. She explains how she learned conservation farming and irrigation techniques to improve the productivity of her plot. Her husband describes how she eventually started her chicken-rearing business, despite five failed attempts, including when she overfed her chickens and they did not lay eggs. She outlines plans to build a greenhouse. She also describes her struggles to succeed at teacher training college, when she was much older than others and struggled to learn, but persisted and achieved good grades. She encourages viewers to continue learning throughout their lives.

Placebo: The placebo videos are about people's lives in rural Western Kenya. They give shots and descriptions of all economic activities which characters undertake in the aspirations video and include any specific information given as part of their stories. They are organised in themes. One section describes people's activities in agriculture and includes all images and information about agriculture from the aspirations video. For example, in the aspirations video, Josefine describes her failure in a chicken farming business. The placebo version has shots and voiceover: "Some people own chickens or have a poultry business. Raising a small batch of chickens can be easy. Raising a big batch is difficult. People who have poultry can also sell eggs. Chickens can be fed, but if they are overfed, they may not lay eggs." Josefine describes learning conservation farming and planning to build a greenhouse. The placebo version shows how some people engage

in conservation farming, which maintains the top soil and preserves the seeds in low rain. Others build greenhouses, which can help grow crops even without rain.

Remaining sections are in similar vein. One shows that some people work for other people as day laborers or teachers, to mirror Josefine's story, where she works as a day laborer and then a teacher. Another describes how people run small businesses, paralleling Judy's story of the progress of her tailoring business and parts of Josefine's story where she weaves baskets. The placebo ensures to cover all specific information conveyed in the aspirations story. For example, describing a tailoring business, it says: "Some tailors think it is important to be careful and creative. They may have business ideas like fixing *mtumba* [second-hand] clothing or sharing a car or weaving a new kind of basket. This can help them make money. Business-people can benefit from certain skills, like calculating their profit on the market. When people work in business, there can be competitors who attract their customers and lower their business. If one business does not do well, sometimes people change to a different business." This section also describes saving for different purposes, including showing savings containers, to parallel Judy and Oyoo's savings practices. A third describes different institutions of learning, to mirror where Josefine trains and works as a teacher, where Judy trains as a tailor and where Judy's son goes to university. Another section describes different ways people live together, to convey any information given by the aspirations video about relationships and marriages. It highlights that people can be married or live alone, can live on their parents' compounds or alone, can fall pregnant, and can have children whom both men and women play with. It also highlights that couples can work together on their farms or in business and can solve problems together.

A.2 Exercise Scripts

Aspirations Workshop:

Field officer (FO): After the participants have watched both the videos: "You have now seen two stories. The first story was about Judy, who decided that instead of accepting her circumstances as being fixed and difficult to alter, she was determined to improve her economic situation for herself and her family. The second story was about Josefine, who decided that it was never too late to learn new skills and that success depends on one's willingness to learn. She continued seeking out opportunities instead of waiting for them to come to her. She found learning and improving to be fulfilling and exciting. Both Judy and Josefine were determined to make plans for their lives and to never give up on these plans. They believed that even people who are from poor backgrounds can make real changes in their own lives using resources that are widely available. Remember, as stated at the start of the video, these stories are parables. The women in the videos are not real people. Their stories are fictional and have been written to teach us lessons. Their stories may not reflect the lives of everyone in Western Kenya."

Best Possible Selves (Drawing): “Now let’s do an exercise like Judy did. Please take a moment to imagine your “best possible self” in 5 years from now. This means that you imagine in 5 years, that everything has gone as well as it possibly could, the best possible way your life could look. You have worked hard and succeeded at accomplishing everything you wanted. Think of this as the realization of your dreams. You have reached your full potential. You can think of what your house will look like, what your farm will look like, what your business will look like, which new abilities you will have developed, or what your family will look like, what your children or grandchildren will be doing. Try to draw what you just pictured. You can draw anything. It does not have to be a beautiful drawing, the only thing that is important is that you know what it shows. The drawing can be as simple as you want, but please try to think of as many details as possible. And please let me know if you need help with the drawing. This is also an exercise that you do on your own, so while you draw, I will ask you to only focus on your own drawing.”

Field officer (FO): While the participants are drawing ask them to explain what they are drawing to keep them engaged. Encourage them and ask for details. E.g. if they draw a house, ask them “How large? What is it made of? Where is it located?” If they draw livestock, ask “Which kind, how many?” If they draw a plot ask “how large? What crop do you want to grow”. Then, for each participant, ask: “Please, explain to the group what your future looks like in as much detail as possible.” Take notes of specific goals for each person. If they give short answers ask clarifying questions as you did while they drew.

Personal Goal-setting: “To achieve your ideal future 5 years from now, it sounds like you should be working towards achieving the following goals: *Field officer (FO): List goals from drawing and make sure to present them in a clear and specific way, e.g. “get a tin roof,” “send children to university”, “get a plot of land”. Exclude things they say that are more “fluffy” such as “know who I am” and “be happy”. List only goals that are concrete and specific: they have to be summarizable in approximately 5 words.*

“You mentioned several goals: [repeat goals from previous round]. Now, I’d like you to think of which of these goals are most important to you and which goals you are most certain that you can achieve by 2021/2. Which is the goal that is the most important for you to achieve that you also think is possible? Which is the second? Which is the third? [Continue until all goals are ranked.]” *Field officer (FO): Write down ranked goals for participant and repeat the list to the participant. Then ask: “How long do you think, realistically, that it will take you to achieve your goal of [insert highest ranked goal] if you actively work towards achieving it already from today?”*

Implementation Intentions: “To achieve your goal of [insert highest ranked goal] you need to have a plan. For this plan to work it has to be as specific and detailed as possible. What

can you do or who can you go talk to today or over the next week in order to work towards achieving your goal of [insert highest ranked goal]?”

Anticipation of Obstacles: “For your plan to work, you need to have a strategy for what to do if something unexpected happens that challenges your plan. For example, you could agree with your family members that you will sit down together and discuss what to do if something unexpected happens that challenges your plan. Or you could identify specific people you will consult with if something happens that challenges your goal. For example, in the first movie Judy saved money in two different places, such that she had one saving that she would only use for achieving her goal and a backup saving she can use if something unexpected happens. What can you do to avoid giving up on your goal if you encounter obstacles?”

Mental Contrasting: “How would your future be better than your life now if you achieved your goal of [insert highest ranked goal]?” “How would it make you feel to achieve this goal?”

“Great work [name of person]! You now know how to set goals and have determined plans and strategies that will help you achieve your most important goals. Remind me again, what are you going to do today or over the next week in order to achieve your most important goal?

Keep the drawing as a reminder of your goals. This can be your life in five years time if you stay focused on achieving your goals and never give up. You can do whatever you want with this drawing but we hope you will put it to good use. For example, you can put it on the wall as a reminder of your goals or to show visitors which goals you are working toward. You can also show it to your spouse or parents to engage them in working towards your goals. Or you could store it in a box or keep it somewhere safe, so that one day you can show it to your children when you have achieved your goal to inspire them to set goals like you have done.

Calendar: As a thank you for your participation today, I’d like to give a wall calendar with the pictures of Judy and Josefine. On this calendar, there is a blank space. First, would you like me to write your most important goal in this blank space?”

“Remember how [Judy and Josefine] overcame obstacles to achieve their goals. Thank you for your participation today and I hope you have found this time we have spent to be useful! I have also brought a bunch of stickers and if you want, you can choose a sticker or multiple stickers that remind you of our conversation, and I would encourage you to place it on your calendar.”

Placebo Workshop:

“You have now seen two short documentaries about Western Kenya. The first video started with a discussion of agriculture in the region: for instance, what food people grow, how they grow it, and what they do with it. It also talked about livestock, like chicken and cows, and agricultural

technologies, like conservation farming. The video then proceeded to talk about marriage, family and children. The second video talked about education: for instance, what skills people can learn, and what institutions they can learn them at. It ended with a discussion about finance and business: for example, what kind of work some people in the region do, what skills their businesses require, and what they do with the money that they earn.”

Field officer (FO): Lead group discussions on these three questions: “Do you think that there was any specific fact that was presented in this video, but was wrong? Or was everything that was said in this movie correct?” “Beyond just the facts, do you think that this movie was well made, or do you think that it is a bad movie? For example, is it rather interesting or rather boring?” “Overall, do you think that this movie accurately describes life in Western Kenya?”

Drawing: “Now let’s do an exercise. I am curious which scene in the movie was most memorable to you. When you think of everything you saw, what is the first thing you remember? Try to draw the scene you just remembered. You can draw anything. It does not have to be a beautiful drawing; the only thing that is important is that you know what it shows. It should be a reminder of what you saw. The drawing can be as simple as you want, but please try to remember as many details as possible. And please let me know if you need help with the drawing. This is also an exercise that you do on your own, so while you draw, I will ask you to only focus on your own drawing.”

Field officer (FO): Let the participants draw for approximately 5 minutes. Ask each participant to explain what she is drawing to keep her engaged. Encourage them and ask for details. E.g. if they draw a house, ask them “How large? What is it made of? Where is it located?” If they draw livestock, ask “Which kind, how many?” If they draw a plot ask “how large? What crop do you want to grow”. Then, for each participant, ask: “Please, explain to the group what your future looks like in as much detail as possible.” Take notes of specific goals for each person. If they give short answers ask clarifying questions as you did while they drew. “Now let’s do another exercise. Think of the last section: the one about business. Different kinds of work were listed in this section: for instance, brick making and tailoring were listed. Some other jobs were also listed. Try to remember one of these jobs. Now please draw a tool that people use when they do this work. You can draw anything. It does not have to be a beautiful drawing; but it should be a reminder of what you saw. The drawing can be as simple as you want, but please try to remember as many details as possible. And please let me know if you need help with the drawing.”

Field officer (FO): Let the participants draw for approximately 5 minutes. Ask each participant to explain what he or she is drawing to keep him or her engaged. Encourage him or her and ask for details. Then, for each participant, ask: “Can you describe what you have drawn in the exercise?”

Calendar: “Feel free to keep the drawing. As a thank you for your participation today,

I'd like to give a wall calendar. I have also brought a bunch of stickers and if you want, you can choose a sticker or multiple stickers that remind you of our conversation, and I would encourage you to place it on your calendar. Thank you for your participation today and I hope you have found this time we have spent together to be useful!"

A.3 Administration of Workshops

The aspirations and placebo workshop administration varied through time for logistical reasons. In a small subset of villages treated early, individuals watched the video and completed the exercises with a single facilitator. In most villages, treatment was in groups of three to four people. We randomly assigned the eligible sample for each village to groups and invited groups to treatment at specific times and locations in their village. People watched the video in pairs on a tablet and completed the exercise together as a group with two facilitators. Those who missed meeting times received the intervention individually. Aspirations and placebo workshop villages within a sublocation all receive treatment either in groups or individually, so treatment implementation is balanced within sublocation. We control for sublocation fixed effects, so aspirations workshops are always administered in the same format as the placebo interventions to which they are compared. As we outline on page 18, group size and composition have no meaningful effect on treatment effects or interpretation of our results.¹

B Departures from Preanalysis Plan

Our estimation and inference methods and our outcome measurement are mostly prespecified. Our preanalysis plan (PAP) is posted at <https://www.socialscisceregistry.org/trials/996>. This appendix describes some departures from the preanalysis plan, none of which substantially changes the estimated treatment effects.

B.1 Measurement

1. We report variables using annual scaling instead of the prespecified monthly scaling to improve comparability to other papers in the cash transfers literature. This multiplies all treatment effects and standard errors by 12, leaving hypothesis test results unchanged.
2. The PAP specifies that the asset aggregate would include net financial liabilities. We omit this component as it is noisily measured.
3. The PAP specifies that household labour earnings (in the revenue aggregate) would include remittances received from household members who are migrants; that the "expenditure on inputs and hired labour" aggregate would include migration expenditures and that the labour

¹The group assignment protocol is described in our pre-analysis plan for data collected straight after the aspirations/placebo intervention, posted at <https://www.socialscisceregistry.org/trials/996>.

supply aggregate would include migrant labour supply. We leave remittances, expenditure and labour of migrants living away from home out of these aggregates as households were uncertain of these values. Migration is relatively rare, with only 4% of labour supply coming from migrants.

4. We use a recall period of the last 30 days for revenue from non-farm enterprises, not the prespecified recall period of the last 30 days that the enterprise was in operation. This increases comparability with all other revenue and investment measures.
5. We rename “investment into economic activity” from the PAP to “expenditure on inputs and hired labour” in order to convey more clearly what is included in this outcome.
6. In the labour supply aggregate, we make two deviations. First, the PAP specified that total labour supply be an index of labour supplied to household agricultural activities, household livestock activities, household non-farm enterprises, and activities outside of the household (casual work, salaried employment, and labour supply by migrants). Instead of the index, we construct a labour supply aggregate measured in days to allow more meaningful interpretation of the magnitude. We scale labour supplied to each activity to annual figures, as outlined in Appendix D. Second, we omit labour supply by household members aged 15 or younger to make the labour supply measure consistent with the prespecified revenue aggregate, which only includes labour earnings from adults.
7. We omit education expenditure from the consumption aggregate because we report education expenditure as part of the education index.
8. We prespecified that we would estimate treatment effects on measures of cognition but dropped these from the endline survey and cannot estimate treatment effects on them.
9. For ease of comparison with the directly elicited aspirations measure, we construct the expectations index using only the directly elicited point estimates for expectations. This index has correlation 0.970 in the endline placebo group with the prespecified index of expectations that uses the means of the individual-specific expectations distributions. So this change makes no difference to our results.

B.2 Econometric Specification

We replace prespecified randomisation block fixed effects with sublocation fixed effects. Sublocations are administrative units in Kenya each containing roughly 10 contiguous villages. The sublocation fixed effects absorb slightly more outcome variation and marginally improve baseline balance. The two sets of fixed effects are highly correlated, with Cramér’s $V = 0.73$.

We prespecified defining the treatment indicators as assigned to receive a cash transfer, assigned to receive an aspirations workshop intervention, and assigned to receive both interventions.

In the paper, we instead use assigned to receive *only* a cash transfer, assigned to receive *only* an aspirations workshop intervention, and assigned to receive both interventions. The respective estimating equations yield identical information from the experiment.

B.3 Summary Indices and Adjustments for Multiple Hypothesis Testing

We depart from the PAP in two ways. First, we adjust for multiple testing by reporting sharpened q -values for the effect of each treatment and for tests of equality of each pair of treatment effects. This approach controls the false discovery rate for tests across different outcomes (Benjamini et al., 2006). We implement this adjustment across all the main economic aggregates and across all the main psychological mechanisms, as well as across the components of each aggregate or index. This is more conservative than the preanalysis plan, which only specified adjusting for multiple testing across components of each economic aggregate. Second, we construct a non-prespecified index, following Anderson (2008), to summarise all the treatment effects on economic aggregates.

B.4 Data Collected Directly After Intervention

We collected some data on aspirations, self-efficacy and growth mindset immediately after the workshops. We call this the “midline” survey. We do not include the midline data in the final paper because we cannot estimate treatment effects that are comparable between the midline and endline for two reasons.

First, the aspirations measures differ enough between survey waves that we cannot compare treatment effects through time: the midline measures cover income and the eldest child’s education, while the endline measures cover assets, income, and the child whose age is closest to 14. (We changed the child identity because qualitative work between the midline and endline showed that the eldest child was more likely to have left education already and parents did not view re-enrollment as plausible.) The self-efficacy and growth mindset questions are comparable between midline and endline but the endline also includes locus of control and combines these three measures into a single index.

Second, comparing treatment effects through time is difficult because the midline and endline measures are observed for different samples. Specifically, 87.2% of the baselined sample completed the endline but 8.5% of the baselined sample completed the endline and missed their assigned workshop, hence missing their midline survey. We do not want to omit these people from the endline analysis because omitting people who do not comply with their randomised workshop assignment may create sample selection bias.

For readers interested in dynamics despite these caveats, we estimate treatment effects on an index of self-efficacy and growth mindset (the measures that are most comparable between midline

and endline) for the 78.7% of the sample who complete both the midline and endline surveys. All three interventions have positive, statistically significant effects on this index at midline that shrink close to zero by endline. But, using the sample of endline respondents, the midline treatment effects of all three interventions have bounds that include zero (Manski, 1990). This suggests the possibility that a temporary rise in self-beliefs may contribute to the treatment effects on economic behaviour. But we cannot rule out the possibility that the positive treatment effects on self-beliefs actually reflect selection bias because we are missing self-beliefs data for part of our sample.

C Details on Cost and Benefit Calculations

This appendix provides additional information about our cost calculations and details on the benefit-cost comparisons. We convert all costs to USD PPP in the analysis month-year (October 2018) to account for inflation and the exchange rates in the same way as the benefit estimates. All costs are per person offered the treatment, not per person who received the treatment, to align with the intention-to-treat benefits estimates.

Panel A of Table C.1 shows three types of costs: the average variable cost of delivering each intervention, the average variable cost of the programme infrastructure and censusing and targeting households, and the fixed costs of delivering the intervention. Together these make up the average total cost of the interventions. Sub-Panel 1 gives the components of average variable cost directly related to delivering the intervention. For the workshop, the intervention delivery costs cover the field operation to deliver the videos and exercises incurred by the implementing partner IPA: field officer training and wages; purchases of tablets, power banks, and projectors (all fully depreciated); and purchases of calendars and stickers for participant exercises. For the cash transfers, the intervention delivery cost is the value of the cash transfer itself. The transfers were delivered by mobile money, so field visits were not required for delivery. The number differs in the cash and combined arms because rates of take-up differed slightly.

Sub-Panel 2 gives programme costs. For cash, the programme costs cover GiveDirectly's field operation to recruit participants, conduct a census, determine eligibility and register eligible households. We also include non-field staff and office costs: a portion of administrative costs to run GiveDirectly in-country and globally, allocated based on this study's proportion of the overall GiveDirectly cash transfer budget. For the workshop, we assume the same costs (i.e. we assume GiveDirectly were running the same census, registration and operation, but delivering the workshop instead of cash). However, GiveDirectly's operational model is unlikely to be the most cost-effective way to deliver the workshop. For example, it might not be cost-effective to target this intervention at poor households, as the costs of a census and then revisiting poor households for registration might be more than the costs of visiting a village once and including all households in the workshop.

Table C.1: Benefit-Cost Comparison of All Interventions

	Workshop	Cash	Combined
Panel A: Costs			
(1) Average variable cost (AVC): intervention delivery	54.3	1,931	2,059
Cash transfer costs	0	1930.8	2005.1
Field personnel and travel	42.3	0	42.3
Intervention materials	5.3	0	5.3
Intervention hardware	6.7	0	6.7
(2) Average variable cost (AVC): programme	218	218	218
Personnel	40.4	40.4	40.4
Field costs and travel	39.8	39.8	39.8
Mobile money and banking	1.4	1.4	1.4
Office	2.7	2.7	2.7
Indirect costs	134.1	134.1	134.1
(3) Fixed costs of intervention development	80.2	0	80.2
Field trip	74.4	0	74.4
Production and editing	5.8	0	5.8
Average total cost (ATC) = (1) + (2) + (3)	353	2,149	2,358
Panel B: Accumulated benefits			
Non-durable consumption ^a	202	267	155
Education expenditure ^a	22.2	44.8	126
Housing and land expenditure ^b	18.0	487	465
Non-land non-housing asset stock at endline ^c	98.0	406	352
Average total benefits	340	1,205	1,098
Panel C: Benefit-cost ratios (BCR) (%)^d			
Average total benefits/ATC	96.4*** (37.6)	56.1*** (6.3)	46.6*** (7.11)
Average total benefits/AVC (delivery only)	627*** (244.31)	62.4*** (7.01)	53.3*** (8.14)
Benefit-cost ratios hypothesis tests (p-value)			
	Cash = Workshop	Comb = Cash	Comb = Workshop
Average total benefits/ATC	.26	.21	.15
Average total benefits/AVC (delivery only)	.02	.29	.02

Notes: We present all figures in USD PPP (October 2018). (a) Accumulated consumption and education expenditure impacts are drawn from intent to treat estimates, using annualised flow impact estimates scaled for each participant by the time between their receipt of the intervention and the endline. Education expenditure is coded as zero for households without school-aged children, while it is coded as missing elsewhere in the paper. (b) This captures household expenditure on land and house assets since the intervention, including repair, maintenance and construction work. (c) Asset stocks are observed in the endline survey. (d) The average total benefits/total cost ratios are the accumulated impacts divided by the cost of each treatment arm. The standard errors on the benefit-cost ratios are the standard errors on the average total benefits divided by the average costs. This leads to substantially higher standard errors for the workshop intervention than the other two interventions, because the substantially higher average costs of the other two interventions mean that the standard errors on their benefits are divided by larger cost denominators.

Sub-Panel 3 gives the costs of intervention development. We set the intervention development cost of the cash transfer to zero because we cannot reliably estimate the cost of creating GiveDirectly. For the workshop, the intervention development cost covers all payments made to Khanga Rue Media, the company that filmed and edited the videos, divided by the number of participants in the workshop and combined arm. This does not include the researchers' time on formative and preparatory work. We exclude the cost of developing the placebo interventions from this calculation.

D Details on Measurement

D.1 Economic Measures

This subsection describes how we construct each of the six main economic aggregates from measures in our survey data. All monetary aggregates are annualised. Any components of aggregates not measured with a 12 month recall are scaled to be annual measures. We convert all component measures into annual scales before combining them into aggregates. Recall periods vary across component measures according to standard measurement practice (Grosh & Glewwe, 2000).

D.1.1 Labour Supply

Total household labour supply is an aggregate of labour supply in the following activities for all household members over 15:

1. **Crop agriculture in last long rains:** Total days worked in the most recent long rains season, multiplied by 1.5 to approximate the annual agricultural labour supply over the short and long rains seasons. The 1.5 scaling factor is based on the baseline survey, where we measured both short and long rains labour supply.
2. **Livestock:** Total days worked in the last 4 weeks for all activities related to livestock raised or owned by the household.
3. **Non-farm enterprises:** Total days worked in the last 4 weeks in household enterprise(s).
4. **Labour outside the household:** In the last 4 weeks, total days worked in casual work or in any salaried employment for someone outside the household.

D.1.2 Expenditure on Inputs and Hired Labour

Expenditure on inputs and hired labour is an aggregate of the following:

1. **Agricultural input expenditures:** An aggregate of total expenditure in the last two seasons on fertiliser, seeds and seedlings, insecticide, storage, farm implements, irrigation and pumping equipment, farm machinery and fuel, and fees or interest for farm-related financial services; total rent paid for plots rented by the household in the last 12 months; and expenditure on hired labour for agriculture. The last is the total number of days of

labour hired in for agriculture in the most recent long rains season multiplied by average wage rates for the village for the activity from the village price surveys, multiplied by 1.5 to approximate the annual agricultural labour demand over the short and long rains seasons.

2. **Livestock input expenditures:** An aggregate of total expenditure on purchasing livestock in the last 12 months; on all livestock inputs used in the last 12 months (feed, veterinary services, medicines and vaccines, equipment, transportation and construction of enclosures); and on hired labour for livestock production. The last is the total number of days of labour hired in for livestock in the last 4 weeks multiplied by wages from the village price surveys. For animals, we only consider production costs in the year of measurement following FAO (1996).
3. **Non-agricultural expenditures:** An aggregate of total expenditure by any non-farming, non-livestock enterprises owned or operated by household members in the last 30 days, plus contributions by the household to any community group economic activities, excluding ROSCAs. Non-farm enterprise expenditures excludes durable goods/assets but includes maintenance (of land, buildings, or equipment), stocks and inventory, wages, fuel and rent costs.

D.1.3 Education index

The education index is an inverse covariance-weighted average, following Anderson (2008), of two variables. For both variables, “school-aged members” are defined as household members aged 6-20. Both variables are set to missing for households with no school-aged members:

- Education expenditure: Total expenditure on education (including school and activity fees, other school related supplies and uniform cost) in the current and preceding school year for all school-aged members, divided by total number of school-aged household members.
- Education participation: Total number of school days attended over the last five days school was in session, divided by the number of school-aged household members. This is set to zero for non-enrolled members and missing for members living in other cities or enrolled in boarding schools.

D.1.4 Revenue

Total revenue is an aggregate of the following::

1. **Revenue from agriculture:** An aggregate of revenue from household crop production and rental income from renting out land across economic activities. Revenue from household crop production is the total value of production for each crop the household grew in each of the last two rainy seasons, including both production sold and production kept and consumed in-kind. We measure units produced by the household. To value the units produced, the calculation uses a hierarchy of valuation approaches, relying on the next best approach

whenever the required data for the preferred approach is not available for a crop.² Where costs have been incurred but output has not been harvested, we include the household estimate of work-in-progress production in output (Grosh & Glewwe, 2000). Rental income is rent received from renting out land or any buildings on it in the last 12 months.

2. **Revenue from livestock rearing and produce:** An aggregate of revenue from livestock sales and livestock production. Livestock sales is the total value of animals the household sold in the last 12 months.³ Revenue from livestock produce is the total value of livestock produce, such as meat, milk or hides, produced by the household in the last 30 days, including produce that is kept and consumed. We use the reported quantity produced and the price per unit produced to estimate the value of production.⁴
3. **Revenue from non-agricultural activities:** An aggregate of total sales, including value of in-kind income, from any non-farming, non-livestock enterprises owned or operated by household members in the last 30 days; total earnings received by any household member from any community group business activity, such as renting out of group assets, in the last 12 months; and earnings in cash or in kind from renting out any assets owned by the household in the last 12 months.
4. **Total household labour earnings:** For all household members in the last 4 weeks, an aggregate of total earnings in cash and in-kind from casual work and total earnings in cash and in-kind from salaried work for someone outside the household.

D.1.5 Non-land Assets

Total household non-land assets is an aggregate of:

1. **Durable assets:** Sum of the respondent’s estimate of the value of household holdings of each asset of different types, if they were to sell them today in their current condition⁵ and

²The hierarchy of approaches to value production is: (a) the respondent’s direct assessment of crop value; (b) the unit price from sales by the same household of the same crop-unit; (c) the crop-unit price obtained from the market price survey; (d) the sub-location median of other households’ direct assessments of the value of the same crop-unit; (e) the sub-location median of other households’ direct assessments of the value of the same crop, converted using a universal unit conversion ratio; (f) the sub-location median of other households’ sales prices of the same crop-unit; (g) the sub-location median of other households’ sales prices of the same crop, converted using the universal unit conversion ratio.

³We unfortunately omitted the value of livestock slaughtered for own consumption at endline.

⁴For livestock products sold by the household, we multiply the quantity sold by the most common price per unit of these sales, which the household reported. For livestock products that are not sold by the household, we multiply the reported quantity with the most common price the household estimated it would receive if it sold its products. If household price data is not available, we value production using local prices, using first prices from the price surveys and then median prices in the sub-location.

⁵This variable sums over the values of two classes of assets in our survey. First, we asked all households to report quantities and values of “core” assets that they own. Second, we asked households to report the value of a randomly selected asset from a list of 8 “randomised” assets, though households reported quantities for all of these assets. For

the respondent’s estimate of the value of any shares household members have in non-financial assets owned by a group.

2. **Livestock:** Total value of the respondent’s estimate of how much money they would get if they sold all mature and immature livestock of different types that they own today.
3. **Savings:** Total value of savings of all household members held at home, with friends and neighbours, with shopkeepers, with microcredit groups, in mobile money accounts, in bank accounts and in ROSCAs.
4. **Stocks of dried maize:** Value of stocks of dried maize currently owned by the household, valued using unit prices of dried maize obtained from the market price survey.

D.1.6 Consumption

Total household consumption is an aggregate of:⁶

1. **Food consumption:** In the last 7 days, value of household consumption of 18 core food items, scaled to capture total consumption following the approach in Beegle et al. (2012) and Nagpal et al. (2024), plus household food consumption outside the home. The PAP details the scaling approach.
2. **Non-food non-durable consumption:** In the last 30 days, value of household consumption of nine core non-food non-durable items, including household goods, fuel, hairdressing and transport fares, scaled to capture total non-food non-durable consumption following the approach in Beegle et al. (2012) and Nagpal et al. (2024). The PAP details the scaling approach.
3. **Expenditure on durable goods:** In the last 12 months, value of household expenditure on durable items and the maintenance of durable items.
4. **Social expenditure:** Value of household expenditure on charitable donations, worship contributions, social and entertainment expenditures, weddings and bride price. Regular worship contributions are 30 days recall, wedding expenditures including bride price are since-intervention recall, and the other items all have 12-month recall periods.

D.2 Psychological Measures

Any indices are constructed as an inverse covariance-weighted average of several components, following Anderson (2008). In the next subsection, we describe how we adapt and validate psychological scales for measures of self-beliefs and depression.

the “randomised” assets without a value, we impute the value based on the reported quantity and the mean value of the asset for the sample (from observations for which we have both variables). The set of “randomised” assets was chosen based on which assets appear least frequently and had low variance in values in the baseline survey.

⁶We do not include actual or imputed value of housing rental expenditure.

D.2.1 Aspirations and Expectations for Future Outcomes

We measure aspirations and expectations in different domains of life – income, assets and children’s education. We define assets as “the worth of your house, your furniture, consumer goods like a TV and fridge and any transport vehicles.” We define income as “all sources of cash income for your household, including what you earn from all agricultural and non-agricultural activities, and money that you have received from any NGO or government programmes.” We measure aspirations and expectations for educational attainment and current educational attainment for one named child, the child aged closest to 14 from the roster of resident household members, following the Indonesian Family Life Survey Round 4. In each of these domains, we measure, in order, the household’s current position (for example, “What is the total value of all the assets owned by your household?”), the respondent’s aspirations, the level the respondent would like their household to reach, and the respondent’s expectations, the level the respondent thinks their household will reach.

The **aspirations index** is an Anderson (2008) index of responses to three questions: aspirations for children’s education from the question “What level of schooling would you like *child name* to achieve?” where *child name* is the name of the child selected from the household roster; aspirations for future assets from the question “What is the level of assets that you would like your household to reach at the end of the next 10 years?”; and aspirations for future income from the question “What is the level of annual income that you would like your household to reach at the end of the next 10 years?”.

The **expectations index** is an Anderson (2008) index of three measures: expectations for children’s education from the question “What level of education do you think *child name* will achieve?” for the same child in the aspirations question; expectations for assets from the question “What is the level of assets that you think your household will reach at the end of the next ten years?” and expectations for income from the question “What is the level of annual income that you think your household will reach at the end of the next ten years?”. We also measure mean expected assets at the end of the next ten years, constructed using the procedure described below; and mean expected annual income at the end of the next ten years, constructed using the procedure described below. These responses are highly correlated with the directly elicited expectations: $\rho=0.868$ for assets and 0.889 for income in the placebo group at the endline. All results in the paper are similar for the directly elicited and estimated measures.

To calculate expectations for household assets and annual income, we elicit probabilistic subjective expectations for assets and income following Dominitz & Manski (1997) and McKenzie et al. (2013) to construct the full distribution of expected assets and income for each respondent. This approach requires four steps, run separately for each of assets and income. First, the enumerator asks

respondents for their estimate of the minimum and maximum value of each variable. Second, based on the minimum and maximum values entered, the survey calculates three respondent-specific thresholds that split the range into four intervals of equal length denoted k_1, k_2, k_3 . Third, the enumerator draws the values of each threshold on a visual aid. They gave each respondent 10 buttons and asked them to allocate buttons to the four intervals, with the number of buttons representing how likely the respondent thinks it is that the realisation of the variable will fall into a given interval. The respondent was required to use all 10 buttons. Finally, these thresholds and interval probabilities allow us to construct $F_{i,k} = P(y_i < Y_{i,k})$ for each respondent i and threshold $k = k_1, k_2, k_3, MAX$. We then fit a respondent-level lognormal distribution and estimate the mean of this distribution.

D.2.2 Expectations about Returns to Specific Investments

The **returns index** is made up of three separate measures of beliefs about returns to investment in different activities. First, we measure beliefs about **the percentage change in yields that will result from use of fertiliser**. Respondents are asked the amount of dry maize they would harvest in the next long rains season from a one acre plot like most other plots in their area, not using fertiliser, and the amount they would harvest when applying 50 kg of DAP (the most commonly used fertiliser locally) per acre while planting. They are told the DAP is free and it is the best, official quality of DAP. Second, we measure beliefs about **the percentage change in yields from more agricultural labour**. Respondents are asked the amount of dry maize they would harvest in the next long rains season from a one acre plot like most other plots in their area, not using fertiliser, and the amount they would earn if working 12 hours more per week. Third, we measure beliefs about the **percentage change in income from investment in university education**. Respondents are asked to estimate the monthly income (at age 30) that their child closest to 14 (as in the aspirations question) would earn if they finished a university degree and the income they would earn if leaving school at the end of secondary schooling (form 4) with a KCSE certificate.

As a rough sense check, we compare respondents' beliefs about returns to econometric estimates from published research on comparable populations. The perceived returns are on average higher but not massively so. We interpret the comparison with caution because of the standard concerns about endogeneity and heterogeneity in returns estimates. The return to fertiliser use belief in the placebo group has median and mean of respectively 114 and 202%, while Duflo et al. (2008) estimate an experimental gross rate of return of 91% to a smaller increase in fertiliser use in maize farming in a nearby part of Kenya. The return to education belief in the placebo group has median and mean of respectively 200 and 322%, while Montenegro & Patrinos (2014) estimate that Kenyans with tertiary education earn on average 122% more than those with secondary education. However, the latter estimate does not take into account the potentially large earnings gains from

education-facilitated rural-to-urban migration, which is a common strategy in this context.

D.2.3 Self-beliefs

The **self-beliefs index** is made of three scales.

1. A Schwarzer & Jerusalem (1995) scale of **self-efficacy**, with high scores indicating high general self-efficacy. Respondents rate the extent to which each of seven statements is true for them from 1 (not at all true) to 4 (completely true). The score is the sum of seven responses, ranging from 7 to 28. Items include “It is easy for me to stick to my aims and accomplish my goals” and “When I am confronted with a problem, I can usually find several solutions”.
2. An adapted version of the Implicit Theories of Intelligence scale (Blackwell et al., 2007) to measure **growth mindset**. The higher the score, the more the respondent has a growth mindset. Respondents are asked to state the extent to which they disagree with each statement on a scale of 1 (agree strongly) to 6 (disagree strongly). The final score is the sum of six responses, ranging from 6 to 36. There are 3 fixed mindset and 3 growth mindset items. We reverse code the 3 growth mindset items. We replace intelligence with skillset as the focus of the measure to make the scale applicable in our context of a rural, adult population. A sample item is “You can always greatly change what skills you have”.
3. A shortened version of the Internal sub-scale from the Internal, Powerful Others and Chance (IPC) scale (Levenson, 1981) to measure **internal locus of control**.⁷ The respondents state the extent to which they agree with each statement on a scale from 1 (disagree strongly) to 4 (agree strongly). The final score is the sum of five responses and ranges from 5 to 20. Items include “When I make plans, I am almost certain to make them work” and “My life is determined by my own actions”.

D.2.4 Preferences

We do two incentivised tasks with the respondents to elicit time and risk preferences. To construct measures of the **discount factor** and an indicator for **present bias**, we use responses from a standard Multiple Price List (MPL) to measure time preferences over money (Coller & Williams, 1999; Andersen et al., 2008). We ask respondents to make a choice seven times between two amounts offered early or later. We do this twice: in the near time frame, we offer money tomorrow or in 15 days and in the future time frame, money in 15 days versus in 29 days. The amount offered at the earlier date is always equal to KSh 400, while the amount offered at the later date increases from KSh 360 to 1600. Among the time questions, respondents are told there is an equal chance for each choice to be randomly selected for payment. We randomize whether the respondents make the

⁷We use 5 of the 8 original items and exclude items not relevant to our population, such as one about driving a car.

decisions in the near or future time frame first and control for the order in which questions appear.

Our measure of the **discount factor** assumes a linear utility function in money and measures the discount factor using the switch from receiving money soon to later.⁸ For example, we ask respondents if they would prefer to receive KSh 400 in 15 days or KSh 440 in 29 days. If they choose KSh 440, then they have a two-week discount factor between 0.91 and 1 and we assign them the mid-point of 0.96. If they choose KSh 400, then we check their choice between KSh 400 in 15 days and KSh 700 in 29 days and so on until they switch. **Present bias** is an indicator for if respondent switches to the (higher) future amount later in the near time frame (tomorrow vs. 15 days), than in the future time frame (15 vs. 29 days). We recode this to a “not present biased” indicator to simplify comparison with other mechanisms.⁹

For **risk preferences**, we use the method developed by Eckel & Grossman (2002). Respondents are asked to make one choice from 6 gambles that are presented to them. Each gamble has two choices: A and B. There is 50% chance of receiving a low payoff (A) and 50% chance of receiving a high payoff (B). The amount the respondent will get for the option they choose will depend on whether A or B is randomly chosen by the computer. One gamble gives a certain return and the other choices increase linearly in risk (as measured by the standard deviation). Under the assumption of constant relative risk aversion (CRRA), the gamble chosen by the respondent corresponds to a coefficient of relative risk aversion and wealth level. This allows for the identification of varying levels of risk aversion. We use the rank of the choices from 1-6, increasing in level of risk. We standardise this measure to have mean zero and standard deviation one in the placebo group, then switch its sign to simplify comparison with other mechanisms. We randomised the order of the certain choice and other choices and control for the order in which questions appear.

Incentivisation works as follows. Before doing the tasks, respondents are told that the computer will draw a lottery across respondents and there is a 1 in 20 chance that they could actually be paid for these tasks. They are told that there is an equal chance for either the time question or risk question to be randomly selected by the computer for payment. Respondents are informed whether they have won anything, the amount won and when they should expect to receive the amount at the end of the survey. All payments are made via M-Pesa. Half the time the time preference task appears first and the other half the risk elicitation task appears first. We control for the order in which the questions appear.

⁸Following Andersen et al. (2008) we use the future time frame for the discount factor to account for any transaction costs or additional risk of future income. We use the first switch point for respondents who switch multiple times.

⁹25% of the people in the placebo group are present-biased over money, consistent with the proportion in other low-income settings: 28% in Ashraf et al. (2006), 17% in John (2020) and 23% in Cassidy (2018).

D.2.5 Mental Health

Mental health is captured with a z-score from a depression severity score calculated from the 10-item Centre for Epidemiological Studies Depression Scale (CES-D) scale (Andresen et al., 1994). Respondents are asked to give the frequency with which they experience symptoms described from 1 (rarely or none of the time, or up to 1 days a week) to 4 (all of the time, or 5-7 days a week). The total score is a sum of all 10 items, with scoring on questions 5 and 8 reversed. We then reverse-code the scale, such that higher scores reflect lower depressive symptoms, to aid comparison with other tests of mechanisms.

D.2.6 Other Measures

1. **Mimicry of videos:** This is the sum of dummy variables coded to one if the respondent engaged in any of the following activities at endline, all of which are featured in the videos: (a) weaved baskets; (b) kept savings in a jar; (c) attended a sewing class; (d) trained as a teacher; (e) grew vegetables to sell on the market.
2. **Field officer-verified assets:** This is the total quantity of the following asset types observed by the field officer: cooking pots and pans, jerry cans, chairs/sofa, tables, radio, TV, poultry house.
3. **Information recall:** An indicator variable if the respondent correctly recalled specific information contained in both videos, about the returns to education for Kenyan men, straight after watching the videos.
4. **Cognitive ability:** An inverse covariance-weighted index of three measures of working memory. These are a digit span measure, where the score is length of the longest sequence that respondents can correctly recall minus two points (Wechsler, 1958); a fluid intelligence measure, the number of correct responses to six Raven’s matrices (Raven, 1990); and cognitive control, measured as the number of correct responses to 3 numerical Stroop tasks comprised of 25 number sequences (Stroop, 1992).

D.3 Development and Validation of Psychological Scales

We translated and backtranslated the psychological scales and aspirations and expectations measures and validated the translations according to standard psychometric practice. Three enumerators did English-Luo forward translation, while another three did blind Luo-English back translation. Differences were reconciled in a group discussion. We then conducted cognitive debriefings with 22 Luo-speaking respondents to capture concepts respondents struggled to understand and identified more context-appropriate concepts to use (Pan et al., 2008). For the psychological scales, we can compare them against standard psychometric evaluation criteria to show scales are

valid and reliable measures (Table D.1). We show results for the endline sample in placebo villages. Results are similar for baseline measures and for the full endline sample in all villages. First, Cronbach’s α values are close to 0.6 or higher for all scales (Column 3 of Table D.1). Cronbach’s α measures the extent to which all the items in each scale measure the same construct, or the inter-relatedness of the items (Cronbach, 1951). Psychologists recommend that scales have α values above 0.7 (Streiner, 2003), although lower values are common with participants with limited literacy (Laajaj & Macours, 2021). Second, all scales meet or almost meet the recommended thresholds for three model fit criteria (Column 4-6 of Table D.1). These measure whether the empirical factor structure in our data matches the structure in the original validation papers.¹⁰ Finally, no item on any scale, or for the aspirations and expectations questions, had non-response rates exceeding 20%, indicating items were well-understood.

The aspirations and expectations measures are not psychological scales, made up of multiple statements with which participants report agreement or disagreement with Likert scale responses. They thus cannot be assessed using these metrics.

Table D.1: Results of psychometric validation

Scale	# obs	Cronbach’s α	Model fit		
			CFI	TLI	RMSEA
Self-efficacy	1753	0.662	0.929	0.894	0.0611
Internal Locus of Control	1756	0.571	0.955	0.909	0.0649
Growth Mindset	1754	0.593	0.955	0.915	0.0651
Depression	1657	0.748	0.946	0.928	0.0529

Notes: This table shows the results of psychometric validation tests. Number of observations varies because we drop observations for which more than 20% of items are missing in a scale.

¹⁰Psychometricians recommend that scales have values above 0.95 for the comparative fit index (CFI), above 0.9 for the Tucker-Lewis index (TLI), and below 0.08 for the root mean squared error approximation (RMSEA) (Cheung & Rensvold, 2002). We use theoretical factor structures from Scholz et al. (2002) for self-efficacy, Levenson (1981) for locus of control, Abd-El-Fattah & Yates (2006) for growth mindset, and Andresen et al. (1994) for depression.

E Additional Results

This supplement shows two additional sets of results that build on analysis in the paper.

E.1 Heterogeneous Treatment Effects

This appendix describes our tests for treatment effect heterogeneity and reports the results. We focus on three outcomes: the index of main economic outcomes (the economic index), the main mechanism outcome (the aspirations index), and one outcome in response to common questions (the mental health index). We estimate treatment effect heterogeneity across the economic index, eight prespecified baseline characteristics – age, the aspirations index, non-land asset value, education, the expectations index, household size, married (versus widowed or unmarried) and the self-beliefs index – and two non-prespecified characteristics in response to common questions – the mental health score and whether the aspirations or placebo intervention was administered in a group.

We estimate heterogeneous treatment effects in two ways. First, we estimate heterogeneous treatment effects using treatment-interacted regressions:

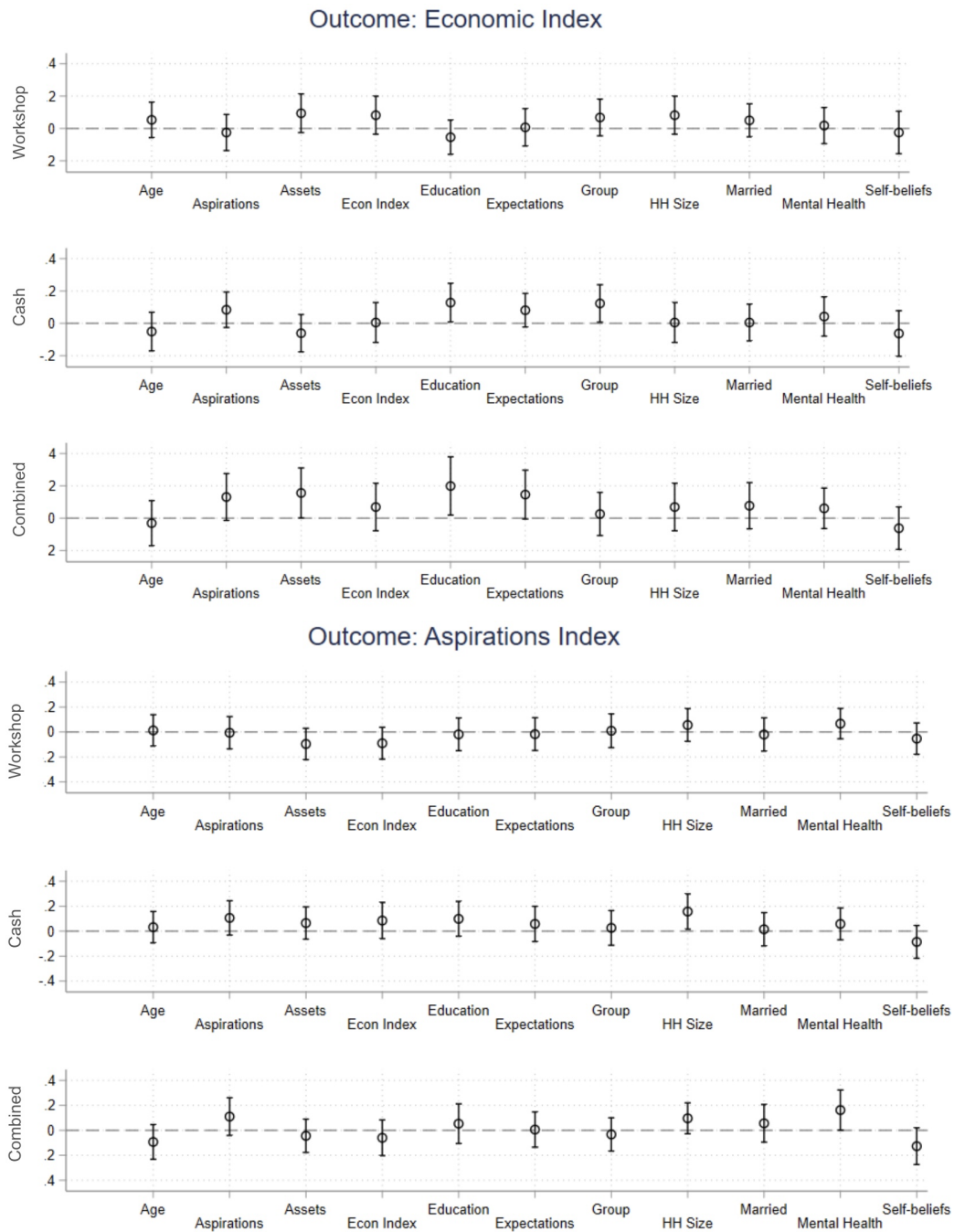
$$Y_{iv} = \text{Cash}_v \cdot \beta_C + \text{Workshop}_v \cdot \beta_P + \text{Combined}_v \cdot \beta_{CP} + \mathbf{X}_{iv} \cdot \mathbf{\Gamma} \\ + \text{Cash}_v \cdot W_{iv} \cdot \alpha_C + \text{Workshop}_v \cdot W_{iv} \cdot \alpha_P + \text{Combined}_v \cdot W_{iv} \cdot \alpha_{CP} + \epsilon_{iv} \tag{1}$$

where W_{iv} is the relevant baseline characteristic. We convert all continuous measures into indicators equal to one for values above the sample median. We display the estimated interaction effects ($\alpha_C, \alpha_P, \alpha_B$) in Figures E.1 and E.2. The interaction effects are seldom large and the fraction of statistically significant estimates is no larger than would arise by chance. These results provide no support for heterogeneity in treatment effects once we adjust for multiple hypothesis testing.

Second, we estimate heterogeneous effects using a causal forest (Wager & Athey, 2018). We first residualise the outcomes with respect to covariates using a standard regression forest.¹¹ We then run a causal forest on these residuals to generate the causal forest estimator. Estimation proceeds as follows. We randomly partition the dataset into training and testing samples in a 80/20 split. In the training dataset, we construct a set of 1001 trees, repeatedly split the data into cells based on values of the nine baseline characteristics and estimate treatment effects within these cells. Each tree is “honestly” fit: the data is used to estimate only the within-leaf treatment effect or to decide on split placement, but not both. We then generate the forest estimate by averaging these prediction rules across trees. Using separate training and testing datasets prevents overfitting. We then apply the causal forest ensemble decision rule to the testing data to estimate heterogeneous treatment effects across the cells.

¹¹Causal forests also require values of the treatment propensity score. We know these exactly from the randomization and hence do not need to estimate them.

Figure E.1: Heterogeneous Treatment Effects Estimated Using Interacted Regression Models

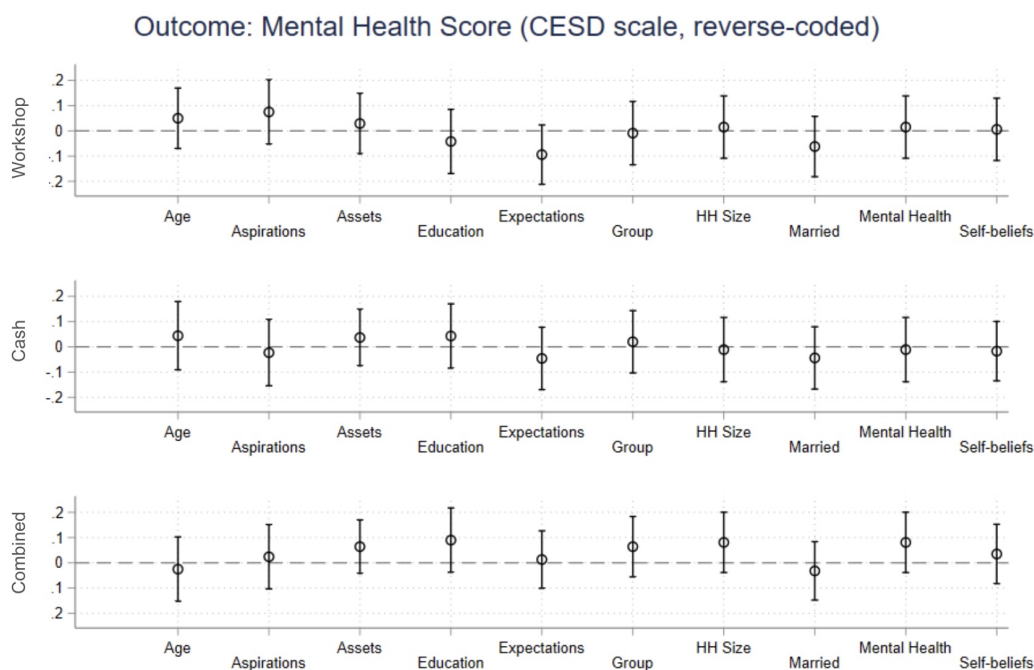


Notes: This figure shows heterogeneous treatment effects estimated using the treatment*covariate interacted regression model in equation 1. The outcome in the first row is the summary economic index, defined in the footnote to Table 2. The outcome in the second row is the aspirations index, defined in the footnote to Figure 1. All variables interacted with treatment are measured at baseline and continuous measures are converted into indicators equal to one for values above the sample median. All variables were prespecified except the economic index; “Group”, which denotes the intervention being administered in groups; and “Mental Health”, which denotes the reverse-coded depression score. Confidence intervals are estimated using heteroskedasticity-robust standard errors, clustered by village. We estimate sharpened q-values to for the false discovery rate across dimensions of heterogeneity, within each row. After this adjustment, no heterogeneous treatment effect is significant at the 10% significance level.

Figure E.3 shows the distribution of heterogeneous treatment effects over the cells for the workshop, cash, and combined interventions. There is limited evidence of heterogeneous treatment effects. The effects of all three treatments on the aspirations index range over the cells from 0 to 0.18 standard deviations, with almost all estimates falling between 0.05 and 0.15. The workshop and combined effects on the economic index are almost all between -0.05 and 0.2 standard deviations, while the cash effects are slightly higher.

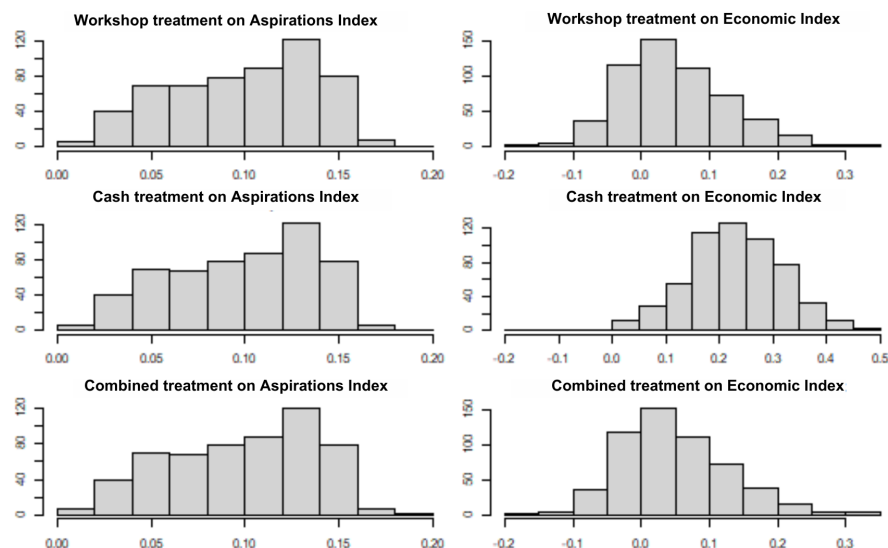
We use these estimates to conduct an omnibus test for the presence of heterogeneous treatment effects, following Tibshirani et al. (2022). Table E.2 shows that we fail to reject the null hypothesis that the treatment effects in all cells estimated by the forest are equal (row 2 of each panel). We also fail to reject goodness-of-fit tests for the forest itself (Table E.2, row 1 of each panel).

Figure E.2: Heterogeneous Treatment Effects on Mental Health Estimated Using Interacted Regression Models



Notes: This figure shows heterogeneous treatment effects estimated using the treatment*covariate interacted regression model in equation 1. The outcome is the Mental Health index, defined in the footnote to Figure 1. All variables interacted with treatment are measured at baseline and continuous measures are converted into indicators equal to one for values above the sample median. All variables except the economic index, “Group”, which denotes the intervention being administered in groups, and “Mental Health”, which denotes the reverse-coded depression score, were prespecified. Confidence intervals are estimated using heteroskedasticity-robust standard errors, clustered by village. We estimate sharpened q-values to for the false discovery rate across dimensions of heterogeneity, within each row. After this adjustment, no heterogeneous treatment effect is significant at the 10% significance level.

Figure E.3: Distribution of Heterogeneous Treatment Effects Estimated Using Random Causal Forests



Notes: This figure shows histograms of heterogeneous treatment effects estimated using the random causal forest. The outcome in the left-hand column is the aspirations index, defined in Figure 1. The outcome in the right-hand column is the summary economic index, defined in Table 2. The x-axis represents the treatment effects.

Table E.2: Tests for Treatment Effect Heterogeneity Using Random Causal Forests

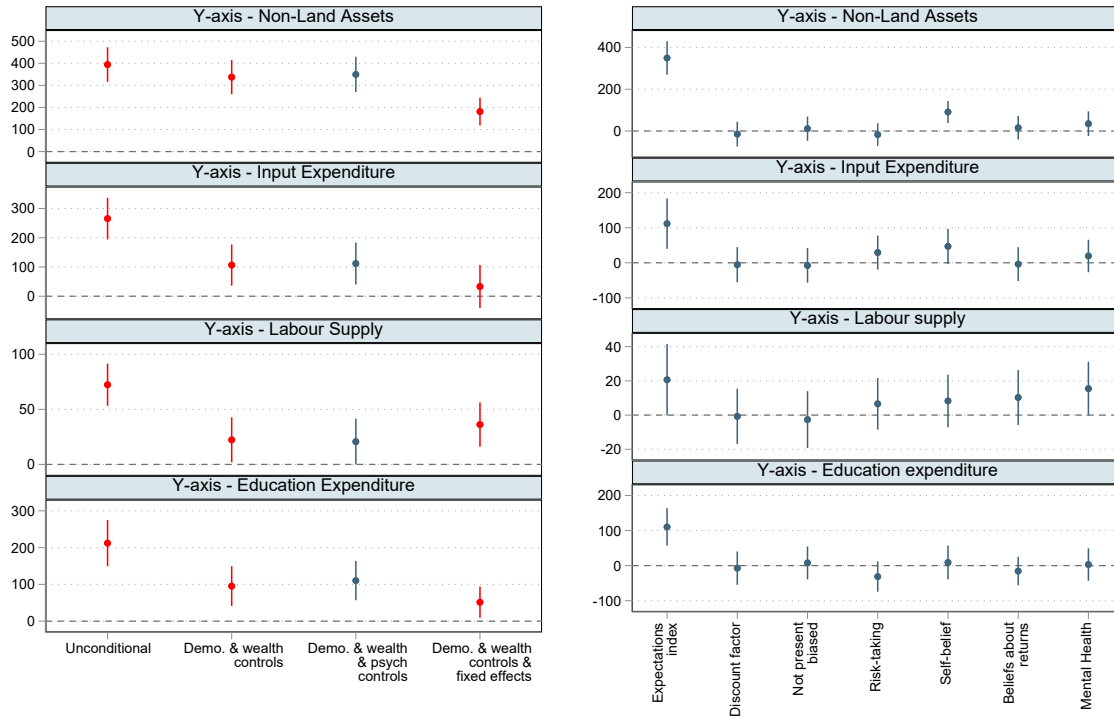
	Workshop	Cash	Combined
Panel A: Economic index			
p-value for null hypothesis: forest fits data	0.98	0.98	0.98
p-value for null hypothesis: forest detects no HTEs	0.82	0.15	0.81
Panel B: Aspirations index			
p-value for null hypothesis: forest fits data	0.96	0.96	0.85
p-value for null hypothesis: forest detects no HTEs	0.64	0.99	0.64

Notes: This table shows results of tests for heterogeneous treatment effects based on random causal forests. All cells display p-values for testing the listed hypotheses. Each column corresponds to a different treatment arm. Each row corresponds to a different null hypothesis. The first row in each panel reports a goodness of fit test; we fail to reject the null hypothesis that the forest estimated using the training data fits testing data. In the second row, we fail to reject the null hypothesis that the effects over all forest cells are jointly equal. The forest uses the same nine baseline dimensions of heterogeneity shown in Figure E.1. The outcome in the first panel is the summary economic index, defined in the footnote to Table 2. The outcome in the second panel is the aspirations index, defined in the footnote to Figure 1. Inference uses a village-clustered bootstrap following Semenova & Chernozhukov (2021).

E.2 Associations Between Wealth, Investment, Expectations, and Other Psychological Characteristics.

Figure E.4 shows the association between wealth, investment, expectations, and other psychological characteristics. This replicates the results in Figures 1 and A.1 replacing the aspirations index with the expectations index, to show that this generates very similar results.

Figure E.4: Relationships Between Wealth, Investment, Expectations, and Other Psychological Characteristics



Notes: This figure shows that, when we replace the standardised aspirations index with a standardised expectations index, the wealth-expectations and investment-expectations relationships are very similar to the wealth-aspirations and investment-aspirations relationships from Section 3.2 and Figure 1.

The four vertically stacked rows in the left panel show coefficients and 95% confidence intervals from regressing different wealth and investment measures on psychological characteristics at baseline. The top row on the left shows results from regressing a wealth proxy (non-land assets) and the second to fourth rows show results from regressing three investment measures (input expenditure, labour supply and education expenditure). Within each left-hand row, the first column shows the coefficients on the expectations index from bivariate regressions; the second column shows the coefficients on the expectations index controlling for respondent age, education, marital status, household size, number of school-aged members, geography fixed effects and (except for the top row) asset value and consumption; the third column shows the coefficients on the expectations index controlling for the same variables and the psychological characteristics shown in the right hand panel; and the fourth column shows the coefficients on the expectations index controlling for the same variables as in the second column, plus respondent fixed effects.

The right panel show coefficients and 95% confidence intervals on all psychological variables from the same regressions that generates the third column in each left-hand row. The right panel is identical to Figure 1 except that we replace the aspirations index with the expectations index.

All asset and investment measures are defined in the footnote below Table 1. The expectations index and psychological measures are defined in Section 3.1. Here, each is standardised to allow for coefficient comparison. All regressions use the endline placebo group data with the top percentiles of expectations, investment, assets, and consumption trimmed. Sample size is 1376 to 1747 depending on the choice of controls and investment measure. The smaller sample sizes are for education expenditure, as this is set to missing for households with no school-aged children. The confidence intervals are estimated using heteroskedasticity-robust standard errors.

Appendix References

- Abd-El-Fattah, S. & G. Yates (2006) "Implicit Theory of Intelligence Scale: Testing for Factorial Invariance and Mean Structure," in *Australian Association for Research in Education Conference, Adelaide, South Australia*, 1–14.
- Andersen, S., G. Harrison, M. Lau & E. Rutström (2008) "Eliciting Risk and Time Preferences," *Econometrica*, 76 (3), 583–618.
- Anderson, M. (2008) "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103 (484), 1481–1495.
- Andresen, E., J. Malmgren, W. Carter & D. Patrick (1994) "Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D," *American Journal of Preventive Medicine*, 10 (2), 77–84.
- Ashraf, N., D. Karlan & W. Yin (2006) "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *The Quarterly Journal of Economics*, 121 (2), 635–672.
- Beegle, K., J. De Weerd, J. Friedman & J. Gibson (2012) "Methods of Household Consumption Measurement Through Surveys: Experimental Results from Tanzania," *Journal of Development Economics*, 98 (1), 3–18.
- Benjamini, Y., A. Krieger & D. Yekutieli (2006) "Adaptive Linear Step-up Procedures that Control the False Discovery Rate," *Biometrika*, 93 (3), 491–507.
- Blackwell, L., K. Trzesniewski & C. Dweck (2007) "Theories of Intelligence and Achievement Across the Junior High School Transition: A Longitudinal Study and an Intervention," *Child Development*, 78 (1), 246–63.
- Cassidy, R. (2018) "Are the Poor So Present-Biased?," *Institute for Fiscal Studies, Working Paper No. 18/24*.
- Cheung, G. & R. Rensvold (2002) "Evaluating Goodness-Of-Fit Indexes for Testing Measurement Invariance," *Structural Equation Modelling*, 9 (2), 233–255.
- Coller, M. & M. Williams (1999) "Eliciting Individual Discount Rates," *Experimental Economics*, 2 (2), 107–127.
- Cronbach, L. (1951) "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16 (3), 297–334.
- Dominitz, J. & C. Manski (1997) "Using Expectations Data to Study Subjective Income Expectations," *Journal of the American Statistical Association*, 92 (439), 855–867.
- Duflo, E., M. Kremer & J. Robinson (2008) "How High Are Rates of Return to Fertiliser? Evidence from Field Experiments in Kenya," *American Economic Review Papers and Proceedings*, 98 (2), 482–488.
- Eckel, C. & P. Grossman (2002) "Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk," *Evolution and Human Behaviour*, 23 (4), 281–295.
- FAO (1996) *A System of Economic Accounts for Food and Agriculture*, Rome: Food and Agriculture Organization.
- Grosh, M. & P. Glewwe (2000) *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*, Washington, D.C.: World Bank.
- John, A. (2020) "When Commitment Fails: Evidence From a Field Experiment," *Management Science*, 66 (2), 503–529.
- Laajaj, R. & K. Macours (2021) "Measuring Skills in Developing Countries," *Journal of Human Resources*, 56 (4), 1254–1295.
- Levenson, H. (1981) "Differentiating Among Internality, Powerful Others, and Chance," in Lefcourt, H. M. ed. *Research with the Locus of Control Construct Vol.1*, 15–63, New York: Academic Press.
- Manski, C. F. (1990) "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80 (2), 319–323.

- McKenzie, D., J. Gibson & S. Stillman (2013) “A Land of Milk and Honey With Streets Paved with Gold: Do Emigrants Have Over-Optimistic Expectations About Incomes Abroad?” *Journal of Development Economics*, 102, 116 – 127.
- Montenegro, C. E. & H. A. Patrinos (2014) “Comparable Estimates of Returns to Schooling Around the World,” *World Bank Policy Research Working Paper*, 7020 (September), 1–41.
- Nagpal, G., K. Orkin & N. Quinn (2024) “Measuring Consumption in 12 Minutes: Optimal Survey Design for Impact Evaluation.”
- Pan, Y., A. Landreth, M. Hinsdale, H. Park & A. Schoua-Glusberg (2008) “Methodology for Cognitive Testing of Translations in Multiple Languages,” in *Survey Methodology* (2008-2), 1–8, U.S. Census Bureau.
- Raven, J. C. (1990) *Standard Progressive Matrices: Sets A, B, C, D & E*: Oxford, UK: Oxford Psychologists Press.
- Scholz, U., B. G. Doña, S. Sud & R. Schwarzer (2002) “Is General Self-Efficacy a Universal Construct? Psychometric Findings from 25 Countries.,” *European Journal of Psychological Assessment*, 18 (3), 242.
- Schwarzer, R. & M. Jerusalem (1995) *Generalised Self-Efficacy Scale*: Windsor, UK: NFER-NELSON.
- Semenova, V. & V. Chernozhukov (2021) “Debiased Machine Learning of Conditional Average Treatment Effects and Other Causal Functions,” *Econometrics Journal*, 24 (2), 264–289.
- Streiner, D. (2003) “Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency,” *Journal of Personality Assessment*, 80 (1), 99–103.
- Stroop, J. R. (1992) “Studies of Interference in Serial Verbal Reactions,” *Journal of Experimental Psychology: General*, 121 (1), 15.
- Tibshirani, J., S. Athey, E. Sverdrup & S. Wager (2022) “Generalized Random Forests Package,” Accessed from <https://grf-labs.github.io/grf/>.
- Wager, S. & S. Athey (2018) “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests,” *Journal of the American Statistical Association*, 113 (523), 1228–1242.
- Wechsler, D. (1958) *The Measurement and Appraisal of Adult Intelligence*, Baltimore, US: Williams & Wilkins Co.