NBER WORKING PAPER SERIES

FEMALE MICRO-ENTREPRENEURSHIP AND CHILD DEVELOPMENT

Andrew Foster
Moritz Poll
Hyun Soo Suh

Working Paper 35710
http://www.nber.org/papers/w35710

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2026

We owe deep gratitude to the participants who dedicated their time to this study; to countless local stakeholders and practitioners who shared their wisdom and advice; and of course to the staff of the implementing partner without whose tireless work this project would not have been possible. This project greatly benefited from discussions with Costas Meghir. Generous funding from the Weiss Fund (BN-002), the National Science Foundation (#2315009), the Dovetail Impact Foundation, the Abdul Latif Jameel Poverty Action Lab (J-PAL GEA-25-03001 & LAI-25-03010 from UK Government International Development funding), the Livelihood Impact Fund, the Gates Foundation (BMG-23-10013-X3 via Innovations for Poverty Action), as well as the Bravo Center, Population Studies and Training Center,Watson Institute, and Nelson Center at Brown University is gratefully acknowledged. Poll also benefited from fellowships funded by the Watson Institute, Brown University Graduate School, the Population Studies and Training Center (under NICHD P2C HD041020 and T32 HD007338), and the Barrett Hazeltine Fellowship for Graduate Research in Entrepreneurship. This study was pre-registered at https://www.socialscienceregistry.org/trials/11789 and approved by Brown University IRB (#98) and Malawi's NCRSH (NCST/RTT/2/6). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Andrew Foster, Moritz Poll, and Hyun Soo Suh. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Female Micro-Entrepreneurship and Child Development
Andrew Foster, Moritz Poll, and Hyun Soo Suh
NBER Working Paper No. 35710
September 2026
JEL No. I24, I25, O12, O15

ABSTRACT

Ultra-poor graduation programs (UPG) reliably reduce extreme poverty among adult recipients, yet their intergenerational effects on children remain largely undocumented. We study a child-focused UPG in rural Malawi that augments the standard graduation bundle with early childhood development (ECD) training, targeting mothers with young children in a population that is both severely stunted and approximately 1 SD behind international benchmarks on cognitive and non-cognitive development. In an RCT of 1,872 households, we show that the intervention raises child development by 0.16 to 0.26 SD within one year but has no significant effects on anthropometrics. Behind the impacts is a broad shift in parental inputs and behaviors. The results demonstrate that a child-focused anti-poverty program can generate child development gains that have not been documented in comparable programs without a child focus.

Andrew Foster
Brown University
Economics Department
and NBER
afoster@brown.edu

Hyun Soo Suh
Washington University in St. Louis
Department of Economics
hyunsoo@wustl.edu

Moritz Poll
Rice University
Department of Economics
moritz.poll@rice.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/11789

1 Introduction

Intergenerational effects are the next frontier in lasting poverty alleviation. Ultra-Poor Graduation programs (UPGs) are some of the most widely studied an implemented anti-poverty programs. They combine cash transfers, entrepreneurship training, financial inclusion, coaching, and a large one-off cash grant or productive assets. UPGs have achieved notable success in reducing extreme poverty lastingly among the targeted parent generation (Banerjee et al., 2015; Bandiera et al., 2017). Yet, an influence of UPGs on child development remains largely undocumented.[1] The broad standardization of household surveys in development economics makes the absence of reported results informative: the standard UPG model likely has at most modest effects on children or else such effects would have been publicized. We show that a specifically child-focused UPG can achieve significant cognitive and non-cognitive child development gains, but no anthropometric impacts in the short run.

We study a program in rural Malawi that targets mothers with young children and augments the standard graduation bundle with an early childhood development (ECD) group training that covers parenting skills, preventive child health, and nutrition. The program does not engage the children directly, which makes parenting behavior the key transmission channel: The business grants and consumption transfers relax the budget constraint, enabling expenditure that may benefit the child. One-on-one coaching and monitoring provides an opportunity for implementers to encourage implementation of learned behaviors in every day life. At the same time, mothers are taught business skills, encouraged to start a micro-enterprise, and join a savings group.

An intergenerational lens highlights an important tradeoff when livelihood programs target young women: Entrepreneurship interventions may raise household income but shift the marginal rate of substitution between time and income. Household income rises but encouraging mothers' labor force participation may adversely impact children (Debela, Gehrke, and Qaim, 2021), especially in contexts with traditionally high childcare burden. However, when mothers control their own hours, their childcare time is relatively inelastic to added market work: mothers protect time with children and absorb the adjustment through other uses of time, not at a cost of child development (Suh, 2026). Since both parental material and time investments are critical inputs in the production of long-term human capital (Cunha, Heckman, and Schennach, 2010), these shifts may generate lasting effects that a purely economic evaluation of the program would miss.

We enrich data from a large randomized control trial in Malawi (Poll, 2026b) with detailed child anthropometrics, cognitive, and non-cognitive development assessments. The intervention randomizes 1,872 households with a child under the age of five at enrollment into a control or treatment arm. Treated households receive a bundled intervention – business and parenting group training with individual coaching, encouragement to form savings groups with peers, regular small cash transfers labeled for consumption especially for children, and a large cash grant labeled for business

[1] Notable exceptions include Brune et al. (2023), who examine long-run UPG impacts on child outcomes in Uganda.

purposes.

Over two years, we collect detailed child development measures alongside parental inputs and characteristics. Anthropometric measures include height-for-age, weight-for-age, and MUAC-for-age, while child development is measured with three instruments for children ages 0 to 9, spanning cognitive and non-cognitive skills. For parental inputs, we construct two latent factors following Heckman, Pinto, and Savelyev (2013) for material investment and parenting quality and collect other rich measures as well as high-frequency time use and expenditure, and direct measures of maternal business and ECD knowledge, schooling aspirations for her children, and mental health.

The intent-to-treat (ITT) effects at the one-year follow-up show significant gains in child development but no effects on anthropometrics. Among children aged 0 to 3, cognitive skills rise by 0.26 SD and non-cognitive skills by 0.24 SD; among children 2 to 9, cognitive skills rise by 0.22 SD. Anthropometric outcomes show no detectable effects, consistent with the shorter time horizon and absence of direct nutrient provision of the program. The effects are not uniform across children. Girls gain somewhat more in terms of their age-adjusted development than boys but the difference is not statistically distinguishable from zero.

The bundled intervention activates a whole host of parental inputs that jointly create child impacts.[2] Material investment in children responds strongly, indicating that the newfound liquidity reaches children, and child-directed expenditure rises across education, health, and food spending. Parenting quality also improves measurably, leading to higher quality of time spent with the child. At the same time, mothers barely reduce the raw time spent with their child. Instead, they crowd in their new business time from reduced casual labor as well as small reductions in idle, leisure, travel, and social time. Because childcare time is close to inelastic with respect to business hours, expanding the business does not force a quantity–quality tradeoff in parental inputs. The opportunity cost falls on the mother's own residual time rather than on the child. Maternal ECD knowledge rises, mothers enroll children in childcare more often, schooling aspirations for children increase, and her mental health and well-being improve. All of these mechanisms have well-documented links to child human capital formation in the literature.

We isolate the cost of the ECD component from the bundled intervention. The Childhoods and Livelihoods program costs approximately $1,350 per household over a two-year period, of which $420 is direct transfers. The ECD component accounts for roughly 13% of the program cost, or about $180 per household over two years and less than $65 per under-five child per year. Adding the ECD component onto a UPG, thus yields a relatively low marginal cost. Consistent with Angelucci, Heath, and Noble (2023), our results highlight how UPGs can achieve specific goals beyond economic empowerment when the design has corresponding programmatic emphasis, and that the marginal cost of doing so is relatively small. While nothing in our study design permits us to assess whether the ECD component is necessary to achieve ECD results, the absence

[2]This study is not set up to investigate the relative contribution of various inputs. Instead we discuss the magnitudes of the impacts on inputs and highlight which channels appear to have little to no movement.

of such impacts in most other UPG evaluations as well as recent work by Bouguen and Dillon (2025) suggest that it might be.

Our work relates to two main strands of literature on graduation programs and early childhood development. First, we contribute to the broad literature on graduation programs. UPGs are widely studied, with over 400 programs worldwide (World Bank, 2024), and consistently improve household consumption, assets, food security, and women's economic outcomes across multiple countries (Banerjee et al., 2015; Bandiera et al., 2017; Balboni et al., 2022). Longer-run evaluations find persistent income gains and evidence of asset accumulation a decade after program completion (Banerjee, Duflo, and Sharma, 2021). Nonetheless, evidence on child-level outcomes is thin. Using the Targeting the Ultra-Poor evaluation, Raza, Van de Poel, and Van Ourti (2018) find gains in acute under-nutrition in Bangladesh but no effect on stunting, consistent with the absence of a child-targeted component. The related multifaceted-program literature points the same way: Bouguen and Dillon (2025) detect anthropometric and ECD gains in Burkina Faso only in the arm that added a direct child-nutrition bundle, not from the cash or asset transfers alone. We address this gap by embedding ECD training in a graduation program and measuring child development across multiple instruments and ages. The Graduating to Resilience program, evaluated by Brune et al. (2023), is the closest to a study showing child impacts of a UPG *without* an ECD component. The program delivers the graduation package to ultra-poor refugee and host communities in Uganda. In the short-run they detect no impacts but a recent six-year follow-up shows signs of ECD impacts. Our program differs by targeting mothers with young children directly and adding ECD training into the graduation bundle (but no direct child stimulation or nutrition assistance). We also assess children using an unusually rich battery of skill assessments and track households over about 40,000 household visits for additional detailed insights.

Second, our work adds onto the rich early childhood development literature. Early parental investments are critical inputs to children's cognitive and non-cognitive skills (Cunha and Heckman, 2007; Cunha, Heckman, and Schennach, 2010), and a large body of experimental work shows that high-quality programs for disadvantaged children translate into persistent skill gains and high economic returns (Heckman et al., 2010; Heckman, Pinto, and Savelyev, 2013; García et al., 2020). Evidence from developing countries supports this: parenting and stimulation interventions in Jamaica (Grantham-McGregor et al., 1991; Walker et al., 2005; Gertler et al., 2014), Colombia (Attanasio et al., 2014, 2020), and China (Zhou et al., 2026) raise language, cognition, and socio-emotional skills. Large-scale preschool and childcare expansions can also improve child outcomes, where effects depend on program quality (Meghir et al., 2023). Most closely related to our design, a set of studies pairs cash transfers with behavioral or parenting components and finds larger early-cognitive gains than from income alone (Macours, Schady, and Vakis, 2012; Akresh et al., 2025). Our study differs from this literature in that the intervention does not interact with the child directly. It reaches children only through changes in maternal income, time, and caregiving, making it a test of whether child-development gains can arise as a byproduct of a livelihood program rather than a purpose-built ECD intervention.

The paper proceeds as follows. Section 2 describes the setting, sample, experimental design, and intervention. Section 3 describes the outcomes of interest and the empirical strategy. Section 4 reports the program evaluation results as well as heterogeneity analysis. Section 5 describes the program cost and Section 6 discusses broader implications and concludes.

2 Child-Focused Ultra-Poor Graduation Program

Extensive details on sampling, treatment assignment, intervention, data collection, and analysis can be found in an online Implementation Appendix.[3] Ethical protections for the vulnerable households in this study, particularly the children, are discussed in a Structured Ethics Appendix, following Asiedu et al. (2021). Below, we focus on the details most pertinent to this paper.

2.1 Setting and Sample

This study took place in Traditional Authority (sub-district) Nankumba (Appendix Figure A1 in Mangochi District in southern Malawi, one of the poorest districts in the country. It targets mothers of children under the age of five whose households live in extreme poverty. To determine program eligibility, a household census of all 28,894 households was conducted, covering a population of 138,000 individuals (Appendix Table A1). Using the Malawi-validated Simple Poverty Score Card (Schreiner, 2019) along with self-reported income, about 10% of households were excluded for not falling under the extreme poverty line of PPP-2022 $2.47 per household member per day. Around 40% of households did not have a child under the age of 5 on the first day of the household census, i.e. a child born on or after 01 August 2018. Another 10% of households were deemed ineligible for not having an adult female or one who was unable to work, or because the data were flagged as implausible or the household did not consent to be surveyed – which was rare. 40% of households were classified as eligible for the program.

Using k-means geographic clustering, roughly 11,000 households were assigned to 271 clusters. Because households tend to live nearby one another, small and remote clusters often correspond to a single village, whereas in larger towns, a village may span several clusters or a single cluster may combine households from more than one village. To determine the sample, 72 clusters were selected into the study in a way that maximizes the distance between clusters and minimizes the potential for spillovers, with buffer clusters in between that are not part of the study.[4] The minimum distance between study clusters corresponds to the distance beyond which Egger et al. (2022) no longer find spillover effects of cash transfers in western Kenya.[5] Appendix Figure A2

[3] Readers may also refer to this appendix for a discussion of the relationship between this and a number of companion papers that focus on the livelihood aspects of this study.

[4] After cluster selection and baseline survey but before the start of the intervention, the local government designated one of our clusters as a flood-prone area and ordered it to be vacated. As our study may have encouraged people to stay in an unsafe area, we shifted to a neighboring cluster in this one instance.

[5] If anything, the study area's road network is substantially less developed, the terrain is sliced up by seasonal rivers

further describes the sampling process.

2.2 Treatment assignment

Clusters were grouped into sets of three ("strata") by proximity that were assigned to a program implementer. In addition, they were randomly assigned to one of three cohorts to be launched a few months apart from each other (see Appendix Figures A1 and A3). The cohorts are timed such that one receives the large cash grant during the planting season, one during the harvest season, and one during the agricultural lean ("hungry") season. Within each implementer's area (stratum), the three clusters were assigned at random to either high treatment saturation (everybody treated), low treatment saturation (half treated, randomly selected at the individual level) or pure control group (nobody treated). Since our focus is on children, we pool data across all three cohorts as well as pooling treated households from high and low saturation clusters into a single treatment group and pure and spillover control groups into a single control group (unless otherwise noted).[6] Stratified random assignment makes the stratum the unit of randomization. All specifications therefore include stratum fixed effects and cluster standard errors at the stratum level (De Chaisemartin and d'Haultfœuille, 2020).[7]

2.3 Intervention

The Childhoods and Livelihoods program consists of five main components (Figure 1), modeled after the ultra-poor graduation program (UPG) originally developed by the NGO BRAC[8] in Bangladesh (Banerjee et al., 2015; Bandiera et al., 2017) with several design modifications discussed

that make effective distances larger than straight-line distances, and personal ownership of bicycles or even motorbikes is lower in southern Malawi than in western Kenya.

[6]The optimal seasonal timing of this program is studied in Poll (2026c). The impact of different levels of saturation on competition between new businesses is analyzed in (Poll, 2026a).

[7]To balance the treatment and control group on a number of key outcome and mechanism variables (see Table C1), the treatment assignment algorithm was run 100,000 times (seeding each with the counter of its iteration). Baseline balance was then estimated under each theoretical treatment assignment and assignments were discarded that lead to a statistically significant imbalance in any of the pre-specified key mechanism variables at the 10% significance threshold (Morgan and Rubin, 2012). The final treatment assignment was then randomly picked among the permissible treatment assignments. This enforces a visually clean balance table. Treatment is assigned separately for each of the three cohorts. As pointed out in both Bruhn and McKenzie (2009) and Morgan and Rubin (2012), classical inference that does not take this rerandomization into account would on average (though not in every case) be too conservative, leading to confidence intervals on the average treatment effect with overly wide coverage and too few rejections of null hypotheses. A Fisher (1935) exact test has the correct coverage, though it tests the sharp "randomization hypothesis" that the treatment had no effect on any one individual rather than no average treatment effect (Ritzwoller, Romano, and Shaikh, 2025). Traditionally, researchers using rerandomization methods have done so somewhat ad hoc, rerandomizing until the balance table is satisfactory. This approach is more transparent. Given $R = 100,000$ simulated assignments and $k = 10$ baseline balance table variables, the likelihood of discarding an assignment at the 10% significance level is $q = 1 - (1 - 0.1)^k$ and the remaining $R \times (1 - q) \approx 34,000$ simulated assignments can therefore be used for the randomization inference (Young, 2019). Table C1 also reports balance on a number of ECD-specific, non-prespecified variables.

[8]BRAC's UPG transfers a productive asset, typically livestock, rather than a large one-off cash grant, paired with training for capitalizing on the asset, and does not involve any child-specific programming.

below that streamline a programmatic focus on children. To the best of our knowledge, this is the first child-focused UPG (World Bank, 2021, 2024).

At the onset of the program, participants are encouraged to form savings groups. For the first six months, treated households receive unconditional monthly cash transfers of $20 (around $75 PPP), labeled for consumption and investments in child health and well-being, as well as for business experimentation. During this period, households also receive a 13-week entrepreneurship soft skills training covering topics such as goal setting, market research, accounting, or microfinance. In parallel, the participants, all of whom are caregivers of at least one child under five, receive a 16-week training course on early childhood development that also covers health topics. Both training courses are taught in groups at the cluster level. After the trainings, participants receive a large, unconditional cash grant of $300, corresponding to about $1,100 PPP[9] and paid in two tranches of $225 at the end of the training and $75 eight months later. The combined cash transfers per treatment household correspond to approximately one year's GDP per capita for the average five-person household in the study area and (because of negative selection into the sample), about 20 months of expenditure of a control group household. Throughout the program's two year duration, households are visited fortnightly for monitoring visits during which treatment households receive one-on-one coaching on the topics covered during the training, particularly preventive health behaviors and ways to avoid or cope with adverse shocks.

The ECD training is a unique component of this UPG design. It spans four domains: (i) child and adult health, including preventive health behaviors, information on and connection to local health services, and guidance on which types of symptoms require immediate medical attention; (ii) child nutrition, including food groups and nutritional diversity; (iii) parenting advice that blends indigenous stories and songs with research on child stimulation; and (iv) children's legal rights to life, healthcare, schooling, a birth certificate, healthy nutrition, play, and expression. Lessons are predominantly taught using example stories, open discussion, and class participation. The curriculum was endorsed by Malawi's Ministry of Gender.

The ECD content is integrated into a graduation program rather than a standalone stimulation curriculum, which distinguishes it from the early childhood parenting intervention on design rather than on quality. The canonical programs are dedicated, age-sequenced curricula, delivered through repeated individualized or group contact at weekly, biweekly, or monthly intervals over 18 to 24 months: e.g. Jamaica (Grantham-McGregor et al., 1991), Colombia (Attanasio et al., 2014), Pakistan

[9]The $300 were converted at the official exchange rate to MWK 530,000 and the World Bank estimates the 2024 PPP conversion factor to be 475 MWK/USD-PPP. The grant was largely unconditional, but the NGO encourages recipients to use it for starting a business or supporting an existing one. To receive the grant, one must attend the training and be able to adequately verbalize a concrete business plan, but participants can miss sessions without penalty and are coached until their plan is satisfactory and the actual implementation of the plan is not afterward enforced. This follows the recommendation in Benhassine et al. (2015). Combined, the seed capital and consumption support total $420 (over $1,500 PPP), which is three times larger than the poverty trap threshold estimated in Balboni et al. (2022), equivalent to the large grants in Haushofer and Shapiro (2016) and marginally lower than those in Egger et al. (2022). The sequencing of monthly consumption support followed by a large lump sum that arrives after sufficient time for deliberation closely matches how poor households themselves report preferring to receive such funds (Kansikas, Mani, and Niehaus, 2023).

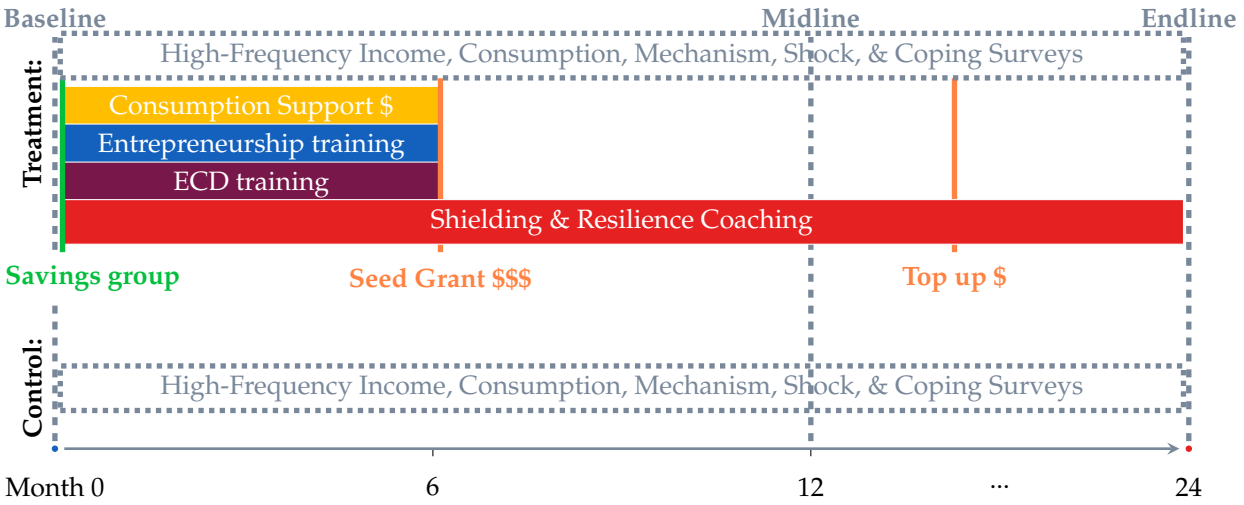


Figure 1: Program Components and Data Collection

Notes: This figure describes the program component sequencing and timing relative to data collection.

(Yousafzai et al., 2014), India (Andrew et al., 2020), and Peru (Araujo, Bosch, and Schady, 2019). The program design instead embeds a four-month group parenting course within the broader intervention and streamlined throughout. Regular coaching visits cover school and childcare center attendance, nutritional diversity, the availability of appropriate clothes, shoes, a blanket, and toys, as well as preventive health practices. The program also monitors children's nutritional status directly.

Both the treatment and control group receive the same logistical and community-level support. Each household is given a basic phone for scheduling surveys and receiving cash transfers or survey incentives through mobile money. Local community-based childcare centers (CBCCs) are supported in-kind with seeds for growing maize and soybeans. CBCC caregivers work on a semi-voluntary basis in exchange for a small stipend and are not generally formally trained. The program provided caregivers with a condensed version of the same ECD training as participants. At the onset of the study, the majority of study clusters had a CBCC, though most of them operate only for a few hours between breakfast and lunch since they do not have the financial means to provide meals to children. Since this support was extended to CBCCs in control group clusters as well, the treatment effects studied here should be interpreted as the marginal impact of the household intervention only.

3 Data Collection, Outcome Measurement and Empirical Strategy

3.1 Data Collection

A team of external enumerators conducted three rounds of main-wave household surveys for the purpose of program evaluation (see Figure 1): As Appendix Figure A3 and the first row in Appendix Table A2 indicate, each cohort started with a baseline survey to all 624 households. A midline survey followed approximately four months after the primary cash grant tranche (about 12 months after the start of the intervention), and an endline 12 months after midline and about 24 months after the start of the intervention (Appendix Figure A3). Main-wave surveys cover a wide range of topics, spanning a household roster and information on livelihoods, health, dwellings, attitudes, and child development outcomes. In addition to these main-wave surveys, all households in the study, both treatment and pure control group, receive monitoring visits from program implementers every two to three weeks.[10] At these visits, implementing staff conduct short surveys about consumption and investment choices, livelihood activities, as well as shocks, coping, and preventive behavior.[11] The treatment group also received one-on-one coaching during these monitoring visits.

3.2 Child Outcomes

The primary outcomes of interest for this study concern early childhood development and include three dimensions: anthropometrics, cognitive, and non-cognitive skills. For anthropometrics, we measure height-for-age (HAZ), weight-for-age (WAZ), and mid–upper arm circumference (MUAC)-for-age (MAZ), for all children that were under age five at enrollment (under age eight by the last endline survey). All measures are referenced against World Health Organization reference tables using the child's age in months and expressed in standard deviations (z-scores).

Stunting in Malawi is severe, yet many development interventions targeting it have not succeeded to move anthropometrics (Manley, Gitter, and Slavchevska, 2013; Manley, Alderman, and Gentilini, 2022). Malawi ranks among the most stunted countries in the world (NSO and ICF, 2017), with 37% of children stunted (Rodriguez-Martinez et al., 2020), and poverty and undiversified nutrition are key drivers of low stature (Black et al., 2013). There is some hope for Cash+ bundles to be more effective (Little et al., 2021) and Bouguen and Dillon (2025) showcased an important proof-of-concept when demonstrating that livelihood interventions in their setting only impacted child anthropometrics when combined with a nutrition provision component. This intervention tests a livelihood-skewed version of that bundle, with 2.5x the size of the cash assistance and a business training, while providing nutritional education rather than direct food assistance.

Timing and evaluation time frames also play an important role here: While acute underweight can be regained in a matter of weeks, height gains typically take time to materialize. There is an

[10] The spillover control group is not visited at high frequency, only for the main wave surveys.
[11] Poll (2026b) investigates these unique high-frequency data in more detail.

ongoing debate over whether the window to achieve height catch-up closes around the 1000-day child age mark (Victora et al., 2010) or not (Leroy et al., 2014).

For child development, we measure cognitive and non-cognitive skills using a battery of age-appropriate assessments for children aged 0 to 9. We combine the Caregiver-Reported Early Development Instrument (CREDI) for ages 0 to 3 — with two assessments directly administered to the child — the Malawi Developmental Assessment Tool (MDAT) for ages 0 to 6 and the short form International Development and Early Learning Assessment (IDELA-SF) for ages 2 to 9. The instruments are complementary where their age ranges overlap: the directly administered MDAT supplements CREDI, which is more prone to experimenter demand and reporting bias (de Quidt, Haushofer, and Roth, 2018). All three have been validated and applied across multiple low- and middle-income country contexts.

The long-form CREDI is designed to be culturally and linguistically neutral, making it well suited for cross-context and low-resource settings. The instrument covers five domains: (i) motor skills (fine and gross motor), (ii) language (receptive and expressive), (iii) cognition (executive function, problem solving and reasoning, and pre-academic knowledge), (iv) socio-emotional development (emotional and behavioral self-regulation, emotional knowledge, and social competence), and (v) mental health (internalizing and externalizing behaviors). Trained enumerators administer it using age- and domain-specific start and stop rules; for example, a 20-month-old child begins in the 18–23 month section, advances to harder items when the mother reports that the child can complete age-appropriate tasks, and stops after five consecutive items the child cannot complete.

We complement CREDI with MDAT (Gladstone et al., 2010).[12] MDAT evaluates child development directly on the child across four domains – gross motor, fine motor, language, and social development – again using direct observation with age-specific start and stop rules. It was developed and validated in Malawi, suiting the context, and serves as a partial validation of caregiver-reported measures. Several items require standardized materials from the MDAT toolkit; for example, fine motor skills are evaluated by asking whether the child can build a tower using three wooden blocks. We trained enumerators in standardized administration protocols covering both theory and practical assessment, working directly with the tool's developer and with trained nurses and psychologists. Enumerators complete a formal assessment and certification on completion, and only those meeting predefined competency standards serve as MDAT assessors.

Lastly, we use short-form IDELA (Pisani, Borisova, and Dowd, 2018). IDELA measures foundational cognitive skills (numeracy, literacy, expressive and receptive language), fine and gross motor coordination, and socio-emotional competencies (self-regulation, prosocial behavior) using ten tasks administered directly on the child.

We try as much as possible to gather repeated measurements of the same child: For anthropometrics, we have the most complete panel data, assessing every child born after 01 August 2018 in every

[12] More details on the MDAT assessments can be found here: https://mdat.org.uk. *Data collection and results for MDAT are in progress and will be included in a future draft.*

survey wave when it is available. Anthropometrics are available for all study waves, while the other child assessment tools were introduced partway through the study: Cohorts 1 and 2 lack CREDI and IDELA at baseline, while cohort 1 lacks CREDI at midline. If a child in a given household was assessed with CREDI or IDELA in a prior main-wave survey and remains within the eligible age range, we attempt to assess the same child again in the next wave. Otherwise, we target the youngest available child within the eligible age range. MDAT is conducted at endline only for all three cohorts, being by far the most logistically demanding and costly of the three measures, but we assess all available children in a given household. Appendix Table A3 summarizes the availability of child outcome measures by cohort and survey wave and Appendix Table A2 shows the sample sizes for each assessment.[13]

For CREDI and MDAT, we benchmark our data using benchmarks provided by the developers. The CREDI benchmarking tool relies on data for children from a broadly pooled international sample of children, meaning that control group means indicate development relative to children globally. MDAT on the other hand uses a sample from Malawi and India for benchmarking, making control group means more a measure of representativeness of children in (severe) poverty. For IDELA, we estimate age-specific control group means and standard deviations non-parametrically, smooth them, and then determine z-scores for all children against that curve.[14]

For each instrument, we construct two summary skill indices – cognitive and non-cognitive – following the conventional scoring approach associated with that instrument. CREDI domain scores are generated upstream by the CREDI scoring app using a two-parameter logistic item response theory (IRT) model (Waldman et al., 2021). We then aggregate these domain scores into a cognitive index (the average of the cognition and language domain z-scores) and a non-cognitive index (the average of the motor and socio-emotional z-scores). For IDELA, we follow Save the Children's official scoring convention (Pisani, Borisova, and Dowd, 2018) and average the relevant task-level subscale z-scores: the cognitive index averages six subscales covering numeracy (number identification, number sensing), literacy (letter identification, oral comprehension), and expressive vocabulary (two tasks), while the non-cognitive index averages four subscales covering emotional awareness and regulation, empathy and perspective-taking, gross motor skills, and fine motor skills. Both indices, for both instruments, are standardized to the control-group mean and standard deviation following Kling, Liebman, and Katz (2007), so that treatment effects are interpretable in units of the contemporaneous control-group distribution.[15]

[13] *Empty columns are survey rounds that have not yet been completed.*

[14] Appendix A.3 provides the details.

[15] Unlike the parental investment indices, we do not fit a confirmatory factor analysis to the child skill outcomes. For CREDI, the upstream IRT-based scoring already reflects item-level factor structure, making a subsequent CFA on four domain composites redundant. For IDELA, equal-weight averaging of subscales is the canonical aggregation approach.

3.3 Parental Inputs

Given the nature of a multi-faceted program, the effect on children could operate through resources, mother's time and interaction with the child, her knowledge, aspirations, or well-being. We therefore track five domains of parental inputs: material inputs, parenting quality, education-related inputs, maternal mental health, and household dynamics. While this does not allow us to assess the relative importance of all inputs, it does provide substantially more nuance about the mechanisms at play: The meaningful sample size allows us to rule out at least some of them as tightly estimated null results.

We construct two latent indices of parental investment from the parenting items collected in each survey wave: *material investment* and *parenting quality*. We specify a confirmatory factor analysis following the measurement-system approach of Heckman, Pinto, and Savelyev (2013), in which each indicator loads on exactly one of the two pre-specified factors, estimate the model on the control group separately by wave, and compute Bartlett factor scores for the full sample.[16] We standardize the resulting scores to the contemporaneous control-group mean and standard deviation following Kling, Liebman, and Katz (2007), so that treatment effects are interpretable in units of the control distribution.[17]

Material investment captures the tangible resources households devote to children and speaks most directly to the cash component. We identify latent material investment from five measures: whether all children own more than three set of clothes, toys, a blanket, a pair of shoes, and ate at least three meals the day before. We supplement the index with raw expenditure data from the high-frequency survey, which records fortnightly records of household food expenditure and child-specific spending on nutrition, health, and education. These data capture both the overall availability of household resources and their allocation towards children, providing a quantitative complement.

Parenting quality captures the non-material investment and speaks to the ECD component. We identify the parenting quality latent factor from seventeen indicators spanning two domains. Educational stimulation covers reading, counting, letter and sound recognition, color and shape identification, hygiene instruction, religious teaching, storytelling, and singing. Parenting practices cover outings, play, chores, physical affection, caregiver satisfaction, discipline, discouraging positive behavior, encouraging eating, and supervision.[18]

Parenting quality can rise because mothers are more knowledgeable and/or their time spent with

[16] Estimating on the control group keeps the loadings free of treatment-induced variation in the investment items; Bartlett scoring yields unbiased estimates of the latent factors.

[17] Results are robust to two alternative indexing approaches: principal component analysis (PCA), which selects weights to maximize total variance explained rather than common variance, and the Kling, Liebman, and Katz (2007) index, which z-scores each item using the control-group mean and standard deviation as baselines and forms an equally-weighted sum within each category.

[18] Negatively oriented items (e.g. whether the child is left unsupervised or is spanked) are reverse-coded so that higher scores indicate higher investment throughout. We also estimate separate latent factors for educational stimulation and parenting practices and results are robust.

the child. We elicit maternal knowledge in three areas – business, ECD, and preventive health – and treat ECD knowledge as the direct target of the parenting training. The high-frequency survey provides fortnightly time use in hours per week across business hours, childcare, household chores, and leisure, which lets us separate a knowledge effect from a time-reallocation effect. The distinction matters because the program raises the opportunity cost of maternal time. Suppose a cash transfer paired with business training pulls mothers toward market work, then the due to the time constraint, any hours the business absorbs must come out of childcare, home production, or leisure. Whether mothers protect time with the child or trade it away for business hours determines whether the intervention faces a quantity–quality tradeoff in maternal inputs. Measuring knowledge and time use jointly lets us distinguish a program that changes what mothers do with a fixed time budget from one that she spends with the child.

Related to child's education, we capture investments that extend beyond the household. We measure whether mothers send their children to community-based childcare centers (CBCCs), which matters in two margins.[19] CBCCs supply early stimulation that substitutes for mother's time, and they relax the childcare constraint that would otherwise bind as mothers expand her business hours. We also elicit mother's expected years of schooling for the youngest child.[20] Expectations are forward-looking investment signals, related to the dynamic complementarity discussed in Foster and Gehrke (2020). If the program were to raise expected schooling without raising current inputs, or the reverse, would imply which constraint binds.

Maternal well-being is another plausible channel as a mother's psychological state shapes the quantity and quality of her interaction with the child. We measure depressive symptoms with the PHQ-9, which asks nine questions about how frequently the respondent experienced depression criterion over the previous two weeks Kroenke, Spitzer, and Williams (2001). We also measure hope with the Herth Hope Index, based on a twelve-item scale spanning temporal orientation towards the future, positivity and expectancy, and interconnectedness with others (Herth, 1992). Both instruments have been widely used and validated and capture two distinct constructs: the PHQ-9 measures current distress, while the Herth index measures optimism, which a growing literature proposes as an input to investment decisions under poverty (Lybbert and Wydick, 2018), though experimental evidence on whether raising aspirations changes behavior is mixed (McKenzie, Mohpal, and Yang, 2022; Orkin et al., 2023).

Lastly, we measure two features of the household environment that may shape how investments translate into child outcomes. We examine women's empowerment along three dimensions: decision-making authority, control over financial resources, and autonomy in child-related decisions. Female agency may itself be a mechanism through which the intervention affects child outcomes, since intra-household bargaining shapes how program resources are allocated (Duflo,

[19] In Malawi, attending school is mandatory from age 6 and in our sample about 90% go to school, so we focus on pre-school attendance.

[20] This question is asked for the youngest son and the youngest daughter who have not yet completed school. For households with both, we average the two responses.

2003). We also examine fathers' involvement in childcare and household activities, which in this context is traditionally low. Father involvement can move in several directions, where he may substitute toward home production as mothers' business hours rise, does not change behaviors, or withdraw if the transfer weakens his perceived obligation to contribute.

3.4 Empirical Strategy

As pre-specified,[21] we evaluate the impact of the UPG program following the ANCOVA specification:

$$y_i = \alpha_0 + \alpha_1 \text{Treatment}_i + \alpha_{BL} y_i^{BL} + \alpha_M \mathbb{1}[y_i^{BL} \text{ missing}] + \alpha_{FE} S_i + \varepsilon_i \quad (1)$$

where the standard errors are clustered at the strata (household cluster triplet) level at which the treatment was assigned. We estimate Equation 1 on various post-treatment outcomes y_i, controlling for baseline y_i^{BL} when available to absorb residual variation in outcomes. When a baseline measure is unavailable for some observations, we include an indicator for missing baseline information. To account for strata randomization, we control for strata fixed effects S_i and we do not control for other baseline characteristics, leveraging the treatment assignment stratification. The coefficient of interest is α_1, which we interpret as the Intent-to-Treat (ITT) effect. All results tables display randomization inference p-values in brackets below the standard errors to account for the way randomization was implemented.

4 Program Evaluation Results

4.1 Sample Balance

Appendix Table C1 demonstrates sample balance across a number of relevant indicators, including child outcomes. At baseline, mothers are on average 31 years old, have themselves had only about 5 to 6 years of education[22] and about 70% of them are literate. Households have on average 3.3 children, 1.4 of which are below the age of 5. There are on average 0.09 pregnancies per household and decisions are made relatively evenly by women and men. Three quarters of houses have a thatch grass roof and food insecurity levels are extremely high. Under-five children are on average 1.4 SD below the WHO benchmark height-for-age (HAZ), 0.8 SD below those for weight-for-age (WAZ), and 0.2 SD below those for mid-upper arm circumference (MUAC) for their age (MAZ).[23] The baseline anthropometric distributions are shown in Appendix Figure A4 and suggests a protein and micro-nutrient deficiency that leads to stunting (reduced height), rather than a caloric deficiency (reduced MUAC) which would be more typical of famine situations. Children are also about 1 SD below the reference values for cognitive, socio-emotional, language, and motor skills as measured

[21] This study was pre-registered at https://www.socialscienceregistry.org/trials/11789.
[22] Malawian primary education is eight years long, meaning that many of them did not complete it.
[23] There is a minor statistically significant imbalance on MUAC but the ANCOVA specification takes care of that.

in the CREDI scale (DAZ). About 41% of children of pre-school age attend pre-school. The factors for material investment and parenting quality are also balanced.

4.2 Impact on Child Outcomes

Table 1 reports the ITT results for child outcomes at midline across two panels: A) anthropometrics, and B) child development. Column (1) reports the control group mean and standard deviation, and column (2) reports the treatment effect from equation (1). These are short-term impacts at midline, roughly one year after the start of the intervention and four months after the main tranche of the cash grant. Column (3) reports the sample size at midline as well as the ANCOVA-relevant sample coverage at baseline.

At the one-year follow up, we do not detect treatment effects for under-five anthropometrics. This is despite the fact that control group children are severely stunted and malnourished, falling below the WHO growth references by -1.40 SD (HAZ), -0.82 SD (WAZ), and -0.38 SD (MAZ), yet the treatment moves none of these measures. The estimates on HAZ, WAZ, and MAZ are all close to and statistically indistinguishable from zero. The estimated standard errors would allow us to reject a null effect for effect sizes of 0.07 to 0.1 SD. In other words, for the oldest children, who are more than six centimeters below international benchmark height, we have enough precision to detect a change in height of half a centimeter but the average height differs by less than a millimeter between treatment and control group children. These are thus very precise short-run null effects on anthropometrics.

The null effects at midline likely reflects the longer horizons needed for nutritional investment to translate into physical growth. This is consistent with a broader literature on income- and cash-based transfers that are not directly targeted at child nutrition (Bouguen and Dillon, 2025): such programs frequently raise food expenditure and intake without moving short-run growth. Our intervention differs in its exact implementation, but the meta-analytic pattern is the same, where the average effect of cash transfers on HAZ is small and statistically insignificant (Manley, Gitter, and Slavchevska, 2013), with the most recent pooled estimate only 0.02 SD and no effect on WAZ (Manley, Alderman, and Gentilini, 2022). In a similar vein, McIntosh and Zeitlin (2024) find no movement in child outcomes within one year in Rwanda; modest improvements in dietary diversity and growth appear only under a much larger lump-sum transfer. Brune et al. (2023) find positive effects when children are exposed to treatment pre-1000 days and find heterogeneous effects.

On child development, the intervention improves ECD outcomes across both instruments and age ranges and for both cognitive and non-cognitive subdomains. Among children aged 0 – 3, assessed with CREDI, children in the control group begin well below the reference norm, at -1.18 SD behind age-adjusted early childhood development (DAZ). The intervention closes part of this gap. The composite score rises by 0.26 SD, with a near identical gain on the cognitive subscale (0.26 SD)

Table 1: Impacts on Child Outcomes

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Panel A: Anthropometrics (age 0–5)			
Height-for-age (HAZ)	−1.390	−0.019	1794
	(1.328)	(0.052)	(✓)
		[0.714]	
Weight-for-age (WAZ)	−0.827	0.035	1802
	(1.128)	(0.045)	(✓)
		[0.445]	
MUAC-for-age (MAZ)	−0.377	0.009	1749
	(0.953)	(0.039)	(✓)
		[0.824]	
Panel B: Early Childhood Development			
CREDI (age 0–3, z)	−1.175	0.263**	520
	(1.363)	(0.110)	(✓)
		[0.029]	
CREDI cognitive (z)	−1.185	0.257**	520
	(1.212)	(0.104)	(✓)
		[0.021]	
CREDI non-cognitive (z)	−1.093	0.236**	520
	(1.221)	(0.104)	(✓)
		[0.032]	
IDELA (age 2–9, z)	0.000	0.159**	1308
	(1.000)	(0.067)	(✓)
		[0.021]	
IDELA cognitive (z)	0.000	0.200**	1308
	(1.000)	(0.075)	(✓)
		[0.012]	
IDELA non-cognitive (z)	0.000	0.097	1308
	(1.000)	(0.062)	(✓)
		[0.125]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

and a smaller, marginally significant gain in non-cognitive skills (0.24 SD). Together, this closes roughly 23% of the baseline deficit. Among children aged 2–9, assessed with IDELA, the composite effect is 0.16 SD, concentrated in the cognitive subscale (0.20 SD), while the non-cognitive effect is marginally statistically significant at 0.10 SD.[24]

Effect sizes are larger for younger children, with CREDI gains exceeding IDELA gains throughout. This age gradient is consistent with the higher malleability of skill formation at younger ages (Cunha and Heckman, 2007) and reinforces the case for prioritizing investment in the first 1,000 days (Black et al., 2017).[25] Point estimates are positive for every development outcome, but the non-cognitive effects are uniformly smaller. These gains are economically meaningful, especially considering that the program is primarily a livelihood intervention with ECD components, but involves no direct child interaction; it reaches them only through the mothers, combining business support with condensed ECD training, and may therefore raise investment through both income and parenting channels.

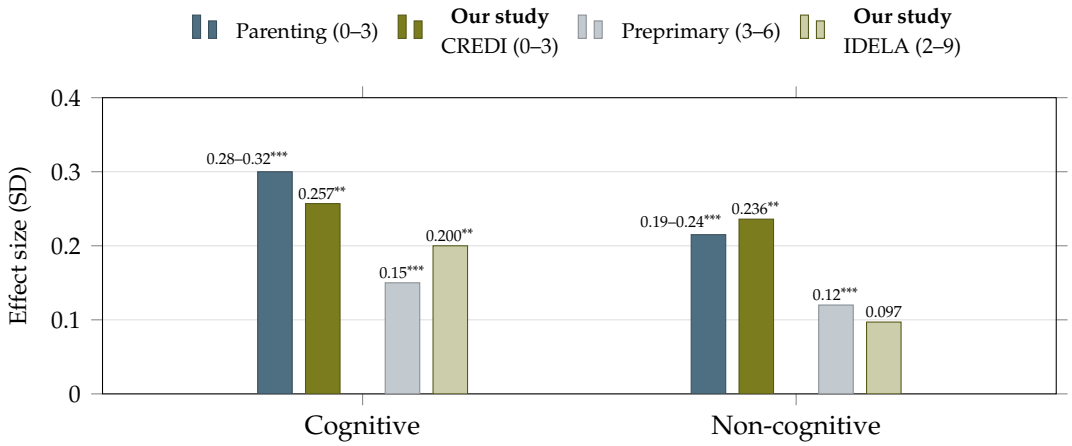


Figure 2: Benchmarking Effect Sizes (SD)

Notes: Bars report standardized treatment effects (in SD) on child development in our study (green) alongside benchmarks from the literature (blue). Cognitive results are shown in the left set of four bars and non-cognitive results on the right. Within each group of bars, effects for younger children are on left and for older children on the right. Our CREDI estimates (ages 0–3) are set against pooled effects of parenting interventions delivered in the first three years of life (Jeong, McCoy, and Yousafzai, 2021); our IDELA estimates (ages 2–9) against estimated returns to preprimary education for ages 3–6 (Holla et al., 2021). For the parenting meta-analysis, the cognitive benchmark spans cognitive and language outcomes, the non-cognitive benchmark corresponds to socio-emotional and motor outcomes. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, Parenting bar height = midpoint of meta-analytic range.

To place the child skill gains in context, Figure 2 benchmarks our development estimates against two meta-analyses of early childhood interventions. Jeong, McCoy, and Yousafzai (2021) synthesize 102 randomized controlled trials of parenting programs delivered in the first three years of life

[24]For this age bracket and scale, no benchmarking data are available, so we cannot assess the relative size of our effects against international benchmarks and normalize them against the control group instead.

[25]At midline, we are not able to rule out that the larger CREDI estimate may be partly induced by experimenter demand effects from the mother's reports. MDAT results are forthcoming for endline to address this concern.

across 33 countries. These include programs designed to improve caregivers' knowledge, practices, and responsiveness and the effect sizes are larger in low- and middle-income countries than in high-income settings. For older ages, we refer to Holla et al. (2021), which pool effect sizes from 55 (quasi-)experimental studies of preprimary education for children aged 3 to 6 across 19 countries.

Against these benchmarks, our effects sit close to the frontier of what child-targeted interventions achieve, even though the program reaches children indirectly through mothers. In the first three years, the CREDI cognitive effect falls closely below the 0.28-0.32 SD range and non-cognitive effect lies within the corresponding range of 0.19-0.24 SD. The fact that such effects emerge from a program centered on business support and condensed ECD training, rather than a dedicated parenting curriculum, suggests the income and parenting channels reinforce one another. For the older cohort, our IDELA cognitive effect is larger than the average return to preprimary education, while the non-cognitive effect is comparable.[26]

4.3 Heterogeneity in Impacts on Child Outcomes

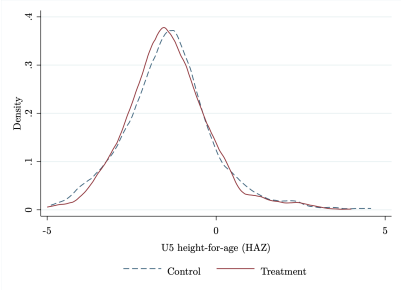
Following the finding of null effects on anthropometrics and positive gains on child cognitive and non-cognitive skills, we compare the raw distribution of child development outcomes across experimental arms, before splitting the sample. Figure 3 plots kernel densities by treatment status at midline for anthropometrics (Panel A) and CREDI and IDELA cognitive and non-cognitive scores (Panel B). While the anthropometric distributions are nearly indistinguishable, distributions of early child skills are shifted slightly to the right for the treatment group in a largely uniform shift that does not appear to have benefited any particular subgroup and is not driven by outliers.

As pre-specified, we look for differential impacts of the program by child gender. Table 2 reports the effects separately for boys (Column 3) and girls (Column 4) and the p-value on the difference (Column 4). There are no anthropometric effects on children of either gender, while treatment improves child development more for girls than for boys. Among younger children, the CREDI composite effect is 0.37 SD for girls against 0.13 SD for boys. The IDELA composite shows the same ordering of 0.21 SD against 0.10 SD. While the effects are statistically significant for girls and not for boys, the gap is not statistically distinguishable from zero. We still interpret the consistently higher and significant impacts on girls as at least modest evidence that the intervention may have somewhat larger effects on girls. It is noteworthy that in this sample, boys and girls start off from roughly the same starting point (Columns 1 and 2), so rather than close a development gap, girls seem to be getting ahead of boys as a result of the intervention. An exploratory random forest approach did not reveal any further insightful sample splits.

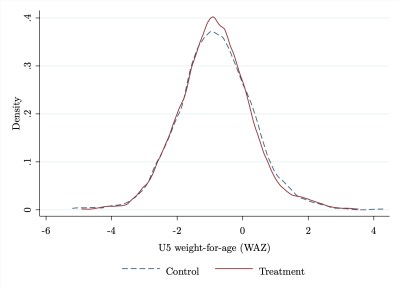
[26] Note that we are benchmarking the one year follow-up results. Effects measured at this stage need not persist. For example, preprimary interventions, combining preschool teacher training with parenting education in Malawi, Özler et al. (2018) find gains at 18 months that fade out by the 36 months follow-up.

Figure 3: Kernel Densities of Child Development Outcomes

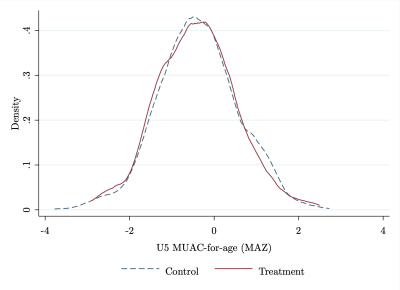
Panel A: Anthropometrics



(a) U5 Height-for-age

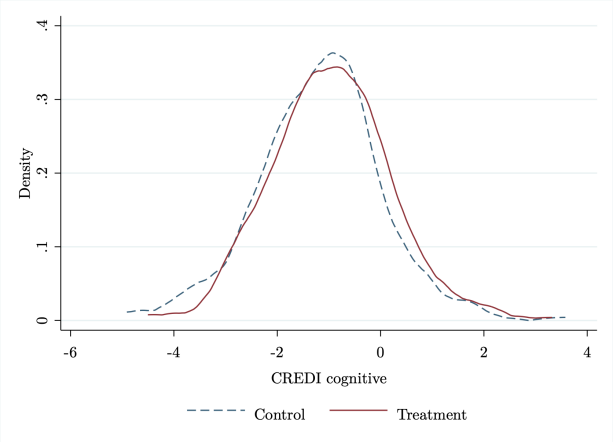


(b) U5 Weight-for-age

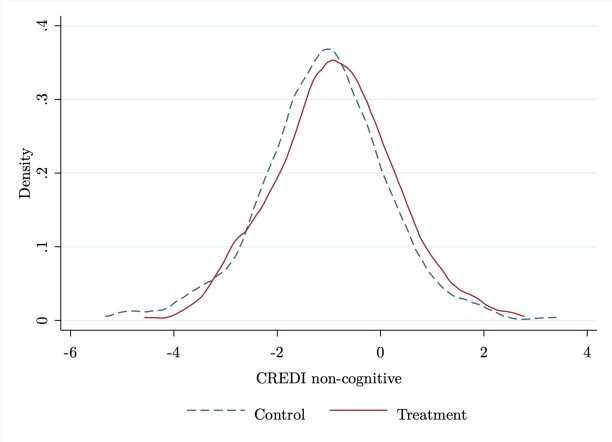


(c) U5 MUAC-for-age

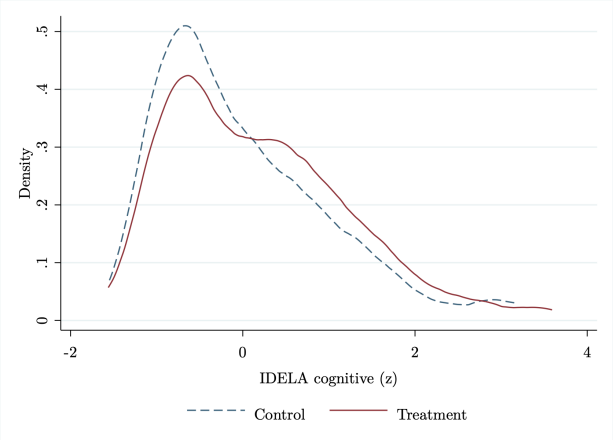
Panel B: Early Childhood Development



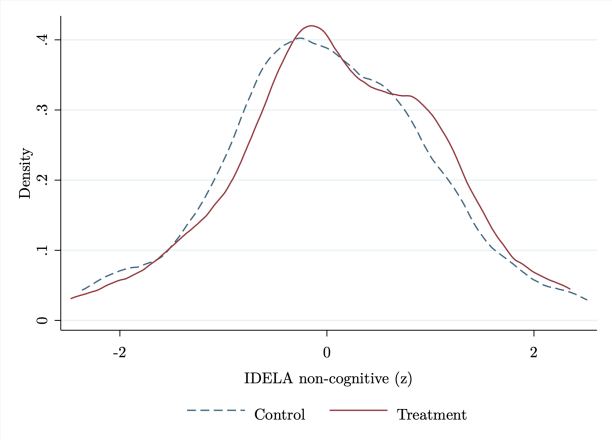
(d) CREDI Cognitive



(e) CREDI Non-Cognitive



(f) IDELA Cognitive



(g) IDELA Non-Cognitive

Notes: Panel A plots the kernel densities of under-five child anthropometric measures and Panel B plots the kernel densities of the score distributions on child cognitive and non-cognitive skills. CREDI is measured for child ages 0-3 and IDELA for child ages 2-9.

Table 2: Treatment Effects by Child Gender

	(1) Control mean (SD) **Male**	(2) Control mean (SD) **Female**	(3) Treatment Effect **Male**	(4) Treatment Effect **Female**	(5) Difference p-value	(6) N (BL)
Panel A: Anthropometrics (age 0–5)						
Height-for-age (HAZ)	−1.454	−1.329	0.024	−0.060	[0.393]	1794
	(1.324)	(1.330)	(0.058)	(0.081)		(✓)
			[0.677]	[0.460]		
Weight-for-age (WAZ)	−0.795	−0.857	0.028	0.044	[0.842]	1802
	(1.168)	(1.087)	(0.048)	(0.069)		(✓)
			[0.587]	[0.511]		
MUAC-for-age (MAZ)	−0.370	−0.382	0.012	0.006	[0.936]	1749
	(0.980)	(0.927)	(0.054)	(0.053)		(✓)
			[0.821]	[0.905]		
Panel B: Early Childhood Development						
CREDI (age 0–3, z)	−1.188	−1.145	0.126	0.365**	[0.353]	519
	(1.483)	(1.242)	(0.165)	(0.167)		(✓)
			[0.466]	[0.043]		
CREDI cognitive (z)	−1.168	−1.184	0.115	0.367**	[0.236]	519
	(1.353)	(1.067)	(0.134)	(0.156)		(✓)
			[0.408]	[0.028]		
CREDI non-cognitive (z)	−1.106	−1.065	0.095	0.344**	[0.244]	519
	(1.329)	(1.110)	(0.152)	(0.142)		(✓)
			[0.542]	[0.029]		
IDELA (age 2–9, z)	−0.028	0.027	0.102	0.209**	[0.358]	1308
	(0.979)	(1.020)	(0.082)	(0.093)		(✓)
			[0.203]	[0.037]		
IDELA cognitive (z)	−0.067	0.064	0.210**	0.187*	[0.846]	1308
	(0.941)	(1.051)	(0.089)	(0.101)		(✓)
			[0.026]	[0.075]		
IDELA non-cognitive (z)	0.000	0.000	0.003	0.182**	[0.115]	1308
	(1.008)	(0.994)	(0.082)	(0.082)		(✓)
			[0.969]	[0.043]		

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.4 Impact on Parental Inputs

To understand the pathways through which the program affects children, we examine its effect on parental inputs. While the bundled nature of the intervention does impact many channels at once, we examine various mechanisms from parental time and money inputs to household dynamics.

Material Inputs The intervention raises the material investment latent factor sharply. Material investment increases by 1.11 SD, relative to a normalized control group mean of zero. This is a large improvement, which is in line with the livelihood support goal of the UPG and the sheer size of transfers relative to control group income.

The material gain is broad but concentrated in durable goods and nutrition (see Appendix Table C2 for the ITT on these factor components). Treatment raises the share of households with safe toys by 33.4 pp (control 28.2%), blanket for every child by 36.1 pp (control 23.1%), and with shoes for every child by 25.1 pp (control 55.5%), and it raises the probability that all children ate three meals the prior day by 4.06 pp (control 36.3%). This pattern is consistent with the lump-sum transfer relaxing the household budget, in which mothers direct resources toward the child-specific durables and food that were previously rationed, while near-universal items such as clothing do not change.

Child-related spending, elicited in monthly monitoring meetings, rises significantly across every category, more than doubling for child-related investment. Treatment increases spending by roughly 40% more on food, 50% more on health, 75% more on education compared to the control mean (Table C3). To benchmark the magnitude, the UPG's monthly consumption support labeled for child well-being, was 34,000 WMK per month and the treated group spent about 9,000 MWK more on children (even after the NGO transfers ceased), which amounts to about a quarter of the support. The high-frequency expenditure data thus confirm that the material investment index reflects a real allocation of resources toward children in both ongoing expenditure and longer-lived consumption items.

Parenting Quality The parenting quality latent factor increases by 0.41 SD.[27] Parenting quality rises by less (in SD terms) than material investment, but the gain is significant and sizable. While the increases in material investment may reflect previously held parental priorities that are enabled by relaxing their budget constraint, parenting quality improvements reflect a change in caregiver behavior. Another reason for the difference in effect size is that while materially these parents are certainly at the very bottom of the distribution, parenting quality measures suggest that many positive practices are widespread even among the control group. We report the item-level estimates in Appendix Table C4.

The quality gain loads on educational stimulation and reduced harsh discipline. Treatment raises the likelihood that parents spent any time in the last week reading to the child by 17.4 pp, teaching

[27] Here, we refer to maternal quality as she received the intervention.

Table 3: Impacts on Parental Inputs

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Panel A: Material inputs			
Material investment (SD)	0.000	1.092***	1828
	(1.000)	(0.057)	✓
		[0.000]	
Child-directed spending (MWK/month)	12,997	9,198***	1529
	(14,319)	(1,242)	×
		[0.000]	
Panel B: Parenting quality			
Parenting quality (SD)	0.000	0.415***	1823
	(1.000)	(0.049)	✓
		[0.000]	
ECD knowledge (0-1)	0.257	0.091***	1829
	(0.099)	(0.008)	✓
		[0.000]	
Childcare & chores (hours in past 7 days)	43.3	-0.941***	1528
	(13.8)	(0.330)	×
		[0.005]	
Panel C: Education			
All U5 children attend CBCC (0/1)	0.433	0.244***	1769
	(0.496)	(0.040)	(✓)
		[0.000]	
Expected child schooling (years)	10.7	0.903***	614
	(2.267)	(0.183)	×
		[0.002]	
Panel D: Mental health			
Mother's PHQ-9 depression score	4.780	-2.808***	614
	(3.419)	(0.325)	×
		[0.000]	
Herth Hope Index (1-5)	3.896	0.207***	1828
	(0.533)	(0.041)	(✓)
		[0.000]	
Panel E: Household dynamics			
Female empowerment (0-1)	0.533	0.009	1367
	(0.246)	(0.013)	(✓)
		[0.490]	
Father involved in childcare (0/1)	0.692	0.021	1782
	(0.462)	(0.025)	(✓)
		[0.403]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

counting by 12.5 pp, teaching letters and sounds by 14.4 pp, and teaching colors and shapes by 15.3 pp. They are more cognitively demanding activities, each off a control mean around half. Disciplinary practices shift in a better direction as well. Mothers are 10.0 pp less likely to spank the child and 11.3 pp more likely to redirect unwanted behavior constructively. Routine behaviors such as hygiene instruction, physical attention, and supervision are unchanged.

We further verify that the self-reported parental investments align with what the enumerators observed during the household survey (Appendix Table C5), covering domains such as hygiene, safe environment, and positive caregiver affect, including questions similar to the Home Observation for Measurement of the Environment (HOME, Caldwell and Bradley, 1984).[28]

We also assess whether mothers recall the core lessons from the ECD training.[29] The results are somewhat nuanced: The control group scores about a quarter of possible points and the treatment group achieves a roughly 9pp higher score during the midline survey (about four months after the cash grant and five months after the last lesson). However, as Appendix Figure C1 shows, this is the product of sharp fade-out and points to the difficulty of reliably assessing this kind of knowledge. Shortly after the trainings end, treated mothers score twice as high as control group mothers. The distinct fade-out likely has to do with how closely curriculum-aligned the assessment is (Cunha et al., 2025). Encouragingly, both reported and observed day-to-day child interaction, as reported previously, remain markedly improved, suggesting that the classroom knowledge has been translated into day-to-day practice.

The experimental design allows us to to investigate an important hypothesized channel behind the success of training and intervention treatments more broadly, knowledge spillovers. The presence of both a pure and a spillover control group enables us to show that immediate neighbors to treated households do not perform any better on the ECD knowledge questions than does the pure control group (Appendix Table C6), highlighting the importance of disseminating information "close to" the child in question. Word of mouth did not appear to spread the training contents beyond the treated mothers.

The amount of raw time spent with the child moves in the opposite direction, but modestly. Treated mothers spend about one hour less per week on home production (childcare and chores, each accounting for about half an hour), consistent with their entry into business drawing some time

[28] Note that effects are on the whole more modest in direct observation (Appendix Table C5) than under self-reports (Appendix Table C4). It is important to remember that this observation took place during the household survey when there was only limited scope for mother-child interaction, which is also evidenced by the much smaller sample size, since children were often not around for enough of the survey to score the interaction. Passive assessments also do not lend themselves to assess the kinds of cognitively stimulating parent-child interactions (reading, story telling, etc.) which would likely drive the cognitive child gains we demonstrate. That said, we cannot fully rule out that self reports are inflated by experimenter demand effects, though even the awareness of what constitutes good parenting practice (i.e. the experimenter's demand) would imply some learning on their part.

[29] For each covered topic, they are asked an open-ended prompts (such as "Name the food groups that you know.", "Name a few behaviors that are recommended during pregnancy.", or "Name the rights of children under 8."). The enumerators have a set of possible correct answers on their tablets that were discussed during group training, which are not shared with the respondent. Instead, respondents give as many answers as they can recall while enumerators check off the correct answers.

away from home. Out of a 100-hour week, this reduction is small. We report detailed time use during the training period of the intervention (i.e. before the cash grant) and after in Appendix Table C7. Both treated and control women devote the largest share of their weekly time into home production, averaging 44 hours per week (about 22 hours each to childcare and chores). Livelihood activities, combining entrepreneurship, casual *ganyu* labor, and farming, amounts to roughly 18-20 hours per week, suggesting women are working part-time.[30]

The share of time allocated to business more than doubles following the cash grant, rising from 7% to 15% of weekly hours. When the program launches, women crowd in the necessary time for both program participation and experimental small businesses from various other activities. When the trainings end, women hold on to their newly carved out livelihood time: the freed up time from trainings is also shifted to the newly started businesses and not back to the previous allocation, suggesting lasting labor supply impacts of the program. Treated women devote about 6.5 hours per week to program involvement, compared to 1.4 hours in the control group (accounting mostly for surveys). When the main training ends, this falls to 3.5 hours in the treatment group and the difference is driven by ongoing coaching visits and attendance to savings group meetings. Compared to the control group's 7 hours of business activities and 8 hours of *ganyu* labor, the treatment group spends 11 hours on business and only 6 hours per week on *ganyu* during the program. After the cash grant was handed out, this gap widened further to over 15 hours of business per week and three hours of *ganyu*. This makes the substitution out of casual agricultural wage labor and into own-account business the sharpest time reallocation for women in our study. Reductions in home production, leisure, and idle time are modest, consistent with evidence that increases in women's market work in low-income settings occur without proportional reductions in domestic responsibilities (Kes and Swaminathan, 2006).

The fact that her increase in business hours is not symmetric to her reduction in home production is an important finding: Women in this context are able to defuse the quantity–quality tradeoff. Mothers instead treat time with the child as a fixed commitment and adjust other margins, consistent with structural evidence that mothers absorb added work out of leisure and home production rather than time with children (Suh, 2026), while parenting quality rises over the same period. This resolves the concern that livelihood intervention targeting mothers pulls attention away from children. At the same time, it points to strong childcare gender norms and a lack of viable alternatives, which relate to the mechanisms discussed below. The inelastic childcare time comes with a cost of leisure to the mother and an inability to pursue either both her business *and ganyu* or allocate even more time to her business to complement the exogenous working capital inflow.

It is also possible that the livelihood choice itself impacts children's development. Both casual agricultural labor and the retail businesses that the majority of women start (Poll, 2026a) are compatible with childcare (Brown, 1970). They allow having the child nearby, can be done near or

[30]Whether travel time should be allocated to business, home production, or leisure is ambiguous and we cannot distinguish whether these trips involve going to the market to buy business inputs, food for the household, or attending social events such as weddings or funerals, so we report it separately.

in the home without a large amount of travel, are not inherently dangerous to the mother or the child, and their productivity does not suffer from the interruptions that come with multi-tasking work and childcare. At the same time, playing in the shade on the shop floor with occasional interaction with a mother waiting for customers is likely much more mentally stimulating to a young child than being strapped to its mother's back under the direct sun on a farm. Beside the obvious income gains, this may also contribute to the more widely documented gap in early childhood development between children of mothers who work in and outside of agriculture (Bliznashka et al., 2023).

Child's Education We observe strong improvements in pre-school childcare enrollment. Roughly half of the control group households' children of the appropriate age attend pre-school and the intervention raises the attendance by 24.5 percentage points (pp), which translate to a 57% increase in take-up. Note that while the intervention included some in-kind and training support to local childcare centers, this support was given to control group clusters as well, so what is isolated here is an encouragement effect to attend, not a quality improvement effect.[31] Relatedly, we ask mothers during regular monitoring visits what they most commonly do with their youngest child when pursuing income-generating activities. Appendix Table C8 shows that most informal childcare mechanisms (like leaving the child alone, with a family member or neighbor, or taking the child with them) decline in favor of leaving them at the childcare center.

Pre-school attendance is widely regarded as important for raising school preparedness. Even though the children in question are themselves not yet of school age, we document that their parents already anticipate being able to keep them in school longer. When asked how many years of education they would like their youngest child to have in an ideal world without cost constraints, parents in both experimental arms quote 12 or more years and the treatment does not change that. When asked, however, how many years they realistically think they can afford to send their children to school, Table 3 shows an increase of roughly one year to the full 12 years of the Malawian school curriculum. A possible channel is therefore the dynamic complementarity between future schooling aspiration and current child skills (Foster and Gehrke, 2020): Parents who expect to be able to maintain the necessary financial stamina to get their children all the way through school may in turn find it optimal to invest more in their children's school preparedness.

Mother's Mental Health Related to aspirations for their children's future, mothers' mental health is also markedly improved by the program. Their PHQ-9 depression score falls by about half (details in Table C9) and they express substantially more hope for their present and future in the 12-item Herth Hope Index (Herth, 1992, details in Table C10). The psychological gains are consistent with the broader literature in which anti-poverty programs improve mental health (Ridley et al., 2020). The program does not target mother's depression, yet better maternal psychological well-being

[31] A full assessment of the quality of local childcare centers is pending but preliminarily, local childcare centers do not seem to suffer from the same quality concerns as described in Özler et al. (2018).

matters for children. In related settings, it raises female empowerment and time- and money-intensive parental investment (Baranov et al., 2020) and lower maternal depression improves mother-child interaction as well as child's own mental health (Cuijpers et al., 2015).

Household Dynamics For household dynamics, we look at female empowerment and father's involvement. The composite female empowerment index does not improve significantly, but the masks heterogeneity across several margins (Table C11). Treated mothers report greater say on how much of household income to save and whether children attend school. The other margins such as meal portioning, child healthcare, discipline, contraception, fertility, and attitudes toward domestic violence are not distinguishable from zero, though the fertility-related ones are negative. We provide more detail on fertility decisions. Table C12 shows no detectable effects: the number of children, children under five, births, pregnancies, and contraception are all small and statistically insignificant, as are parents' stated ideal number of children. It is worth noting, though, that both the negative coefficients on female agency over fertility and the positive coefficient on pregnancies are very close to conventional statistical significance thresholds.

At the same time, the program does not induce fathers to partake more in child rearing. It is worth remembering that the assessed intervention does not directly address fathers. Early results from a recent evaluation of a couple-focused UPG intervention by Bedi et al. (n.d) in nearby TA Makanjira looked more promising on that front.

5 Program Cost

Given that the intervention is a multi-faceted graduation program bundling transfers, training, and coaching, the cost of any single component is difficult to isolate. In this section, we discuss the necessary assumptions for identifying and benchmarking the overall costs of the program, before repeating the process for the ECD component in isolation.

Overall Program Cost This section follows Dhaliwal et al. (2013) in assessing the cost per served household. In doing so, the following assumptions were made. Cost estimates represent the total cost, i.e. a blend of fixed and marginal costs. This makes them an upper bound for the marginal cost around the current program scale of 936 treated households. Adjustments were made to account for the costs of research that would not be incurred in regular programming. For example, each RCT cohort was implemented by eight field staff and one supervisor but without a need for surveying the control group, six field staff and one supervisor would have sufficed. Similarly, the RCT eligibility census was done at a much larger scale in order to create buffer room in between clusters, which would not be necessary in regular operation. Since the implementing partner is US-funded, local inflation and exchange rate arithmetic can be sidestepped by compiling costs on an accrual basis in USD that were transferred to Malawi. This does at the same time highlight one of the major

cost drivers of the program: The Malawian government exchange rate regime relies on a mandated exchange rate that is set at about twice the true value of the local currency, as measured by the black market rate. Besides easing imports for the few with access to these artificially cheap dollars, this highly controversial policy severely hampers exports, leads to crippling foreign exchange shortages, and makes implementation of development and humanitarian programs about twice as expensive in Malawi than in its neighboring countries.

The total cost per household of the Childhoods and Livelihoods program is $1,350 over a two-year period, which includes $420 of direct transfers. As is typically done, this estimate does not include any indirect costs.[32] Costs reflect spending from 2023 to 2026-H1 while costs for 2026-H2 and 2027 are projections. This puts the evaluated program at the upper end of the documented range of $400 to $1,700 (World Bank, 2024), though absent the exchange rate cost and considering the large share of direct cash transfers and its two-year duration, it is fairly representative.

Cost of the ECD component Conducting a four-month standalone ECD training course for extremely poor Malawian households and then monitoring and coaching them for the remainder of two years would likely be a relatively costly endeavor. Including child-focused programing into an existing UPG, however, is comparatively cheap. In the case of the evaluated program and with a two-year UPG in place, our best estimate is that the ECD components make up about $180 of the above-mentioned $1,350 (and the same exchange rate arithmetic as above applies): Implementers would save only about two hours per week, enough to increase the number of participants they monitor by 18.5%. This is equivalent to $130 per under-five child or $65 per child per year – a potentially attractive cost for any UPG implementer looking to add an early childhood component, compared to others in the literature: For example, the Jamaica Reach Up and Learn and China REACH programs cost $750 and $525 annually per child (Zhou et al., 2023), though they do achieve larger effects.

6 Discussion and Conclusion

This paper examines whether a child-focused ultra-poor graduation program improves child outcomes – anthropometrics, cognitive, and non-cognitive child development – and through which parental inputs. At the one year follow-up, the intervention had no effects on anthropometrics but promising gains in early childhood development. The intervention generates significant gains in both cognitive and non-cognitive skills for children ages 0 to 9, and bigger effects for younger

[32]Three scenarios for indirect costs would seem appropriate. The first is to ignore them, as do most implementers. The second is a fixed proportion of 10-20%. The last would be to split actual overhead across all programs implemented by the NGO. The latter, while closest to the truth, is at the same time least informative for future programming: The evaluated program was launched before and ended after the shutdown of USAID, which in turn triggered a severe funding cut for humanitarian and development projects around the globe which all of the other programs that the implementer was conducting at the time fell prey to. A similar overhead would therefore be split across many times more households in the first half of the project than in the second half.

children. The bundled intervention pushes along several parental input margins, ranging from material inputs and parenting quality to childcare enrollment, educational aspirations, parenting knowledge, and mental health. All of these are in addition to the typical livelihood gains from a graduation program (Poll, 2026b).

We find no evidence of an adverse time trade-off between childcare and mothers' livelihood engagement. Rather, hours are crowded in from a variety of other activities to make room for new businesses, while keeping childcare responsibilities largely unchanged. This pattern of part-time self-employment is similar to the broader context in Sub-Saharan African countries, where female labor force participation is high but average market hours remain low (Dinkelman and Ngai, 2022). Shifting from casual agricultural labor to petty trading (Poll, 2026a) highlights that women choose their livelihoods with childcare compatibility in mind (Brown, 1970) and the new businesses may allow scope for multi-tasking and child stimulation that working in a field does not.

A key open question is whether ECD programming is necessary to achieve ECD results. This study is not set up to answer that question and there is genuine equipoise in the literature. To our knowledge, apart from Bouguen and Dillon (2025) and Brune et al. (2023), no other broad livelihood support program shows short-term improvements in early childhood development and in their paper, only treatment arms that included a dedicated nutritional component affected children. While for many studies, such outcomes were not measured (e.g. Banerjee et al., 2015, to the best of our knowledge did not), those that did found no impacts or they materialized only years later in program designs without child targeting (cf. Tamim et al., 2026, or a study in Uganda that does not yet have a citation). For a reader who is inclined to believe that ECD programming is necessary to achieve ECD impacts, about $65 per under-five child per year is very good value compared to other interventions in the literature.

References

Akresh, Richard, Damien de Walque, Harounan Kazianga, and Abigail Stocker. 2025. "Medium-Term Impacts of Integrated Social Safety Nets: Cash Transfers, Information Meetings, and Home Visits for Child Development." Tech. rep., National Bureau of Economic Research.

Andrew, Alison, Orazio Attanasio, Britta Augsburg, Monimalika Day, Sally Grantham-McGregor, Costas Meghir, Fardina Mehrin, Smriti Pahwa, and Marta Rubio-Codina. 2020. "Effects of a scalable home-visiting intervention on child development in slums of urban India: evidence from a randomised controlled trial." Journal of Child Psychology and Psychiatry 61 (6):644–652.

Angelucci, Manuela, Rachel Heath, and Eva Noble. 2023. "Multifaceted programs targeting women in fragile settings: Evidence from the Democratic Republic of Congo." Journal of Development Economics 164:103146.

Araujo, Maria Caridad, Mariano Bosch, and Norbert Schady. 2019. "Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap?" The Economics of Poverty Traps :357–382.

Asiedu, Edward, Dean Karlan, Monica Lambon-Quayefio, and Christopher Udry. 2021. "A call for structured ethics appendices in social science papers." Proceedings of the National Academy of Sciences 118 (29):e2024570118.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2020. "Estimating the production function for human capital: results from a randomized controlled trial in Colombia." American Economic Review 110 (1):48–85.

Attanasio, Orazio P, Camila Fernández, Emla OA Fitzsimons, Sally M Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. 2014. "Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial." BMJ 349.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil. 2022. "Why do people stay poor?" The Quarterly Journal of Economics 137 (2):785–844.

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

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." Science 348 (6236):1260799.

Banerjee, Abhijit, Esther Duflo, and Garima Sharma. 2021. "Long-term effects of the targeting the ultra poor program." American Economic Review: Insights 3 (4):471–86.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial." American Economic Review 110 (3):824–859.

Bedi, Tara, James Copestake, Michael King, and Julia Vaillant. n.d. "Female Empowerment, Well Being and Marriage Quality: A Mixed-Method Evaluation of a Gender Focused Multifaceted Anti-Poverty Programme in Malawi." URL https://www.tcd.ie/time/research/malawigraduation/. Working paper.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2015. "Turning a shove into a nudge? A "labeled cash transfer" for education." American Economic Journal: Economic Policy 7 (3):86–125.

Black, Maureen M, Susan P Walker, Lia CH Fernald, Christopher T Andersen, Ann M DiGirolamo, Chunling Lu, Dana C McCoy, Günther Fink, Yusra R Shawar, Jeremy Shiffman et al. 2017. "Early childhood development coming of age: Science through the life course." The Lancet 389 (10064):77–90.

Black, Robert E., Cesar G. Victora, Susan P. Walker, Zulfiqar A. Bhutta, Parul Christian, Mercedes de Onis, Majid Ezzati, Sally Grantham-McGregor, Joanne Katz, Reynaldo Martorell, and Ricardo Uauy. 2013. "Maternal and Child Undernutrition and Overweight in Low-Income and Middle-Income Countries." The Lancet 382 (9890):427–451.

Bliznashka, Lilia, Joshua Jeong, Lindsay M. Jaacks, and Prashanth Nuggehalli Srinivas. 2023. "Maternal and Paternal Employment in Agriculture and Early Childhood Development: A Cross-Sectional Analysis of Demographic and Health Survey Data." PLOS Global Public Health 3 (1):e0001116.

Bouguen, Adrien and Andrew Dillon. 2025. "Beyond poverty reduction: Evidence from a multifaceted program on poverty, nutrition and child development." Journal of Development Economics :103596.

Brown, Judith K. 1970. "A Note on the Division of Labor by Sex." American Anthropologist 72 (5):1073–1078.

Bruhn, Miriam and David McKenzie. 2009. "In pursuit of balance: Randomization in practice in development field experiments." American Economic Journal: Applied Economics 1 (4):200–232.

Brune, Lasse, Nathanael Goldberg, Dean Karlan, Doug Parkerson, and Christopher Udry. 2023. "The Impact of a Graduation Program on Livelihoods in Refugee and Host Communities in Uganda." Research study, Innovations for Poverty Action. URL https://poverty-action.org/impact-graduation-program-livelihoods-refugee-and-host-communities-uganda.

Caldwell, Bettye M. and Robert H. Bradley. 1984. Home Observation for Measurement of the Environment. Little Rock, AR: University of Arkansas at Little Rock.

Cuijpers, Pim, Erica Weitz, Eirini Karyotaki, Judy Garber, and Gerhard Andersson. 2015. "The effects of psychological treatment of maternal depression on children and parental functioning: a meta-analysis." European Child & Adolescent Psychiatry 24 (2):237–245.

Cunha, Flavio and James J. Heckman. 2007. "The Technology of Skill Formation." American Economic Review 97 (2):31–47.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." Econometrica 78 (3):883–931.

Cunha, Flavio, Qinyou Hu, Andrea Salvati, Kenneth Wolpin, and Rui Zeng. 2025. "An evaluation of the Alief Independent School District JumpStart program: using a model to recover mechanisms from an RCT." Tech. rep., National Bureau of Economic Research.

De Chaisemartin, Clément and Xavier d'Haultfœuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." American Economic Review 110 (9):2964–2996.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." American Economic Review 108 (11):3266–3302.

Debela, Bethelhem Legesse, Esther Gehrke, and Matin Qaim. 2021. "Links Between Maternal Employment and Child Nutrition in Rural Tanzania." American Journal of Agricultural Economics 103 (3):812–830.

Dhaliwal, Iqbal, Esther Duflo, Rachel Glennerster, and Caitlin Tulloch. 2013. "Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education." Education policy in developing countries 17:285–338.

Dinkelman, Taryn and L Rachel Ngai. 2022. "Time use and gender in Africa in times of structural transformation." Journal of Economic Perspectives 36 (1):57–80.

Duflo, Esther. 2003. "Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa." The World Bank Economic Review 17 (1):1–25.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. 2022. "General equilibrium effects of cash transfers: Experimental evidence from Kenya." Econometrica 90 (6):2603–2643.

Fisher, Ronald Aylmer. 1935. The design of experiments. Edinburgh: Oliver and Boyd.

Foster, Andrew D and Esther Gehrke. 2020. "Start what you finish! Ex ante risk and schooling investments in the presence of dynamic complementarities." Tech. rep., National Bureau of Economic Research.

García, Jorge Luis, James J Heckman, Duncan Ermini Leaf, and María José Prados. 2020. "Quantifying the life-cycle benefits of an influential early-childhood program." Journal of Political Economy 128 (7):2502–2541.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M Chang, and Sally Grantham-McGregor. 2014. "Labor market returns to an early childhood stimulation intervention in Jamaica." Science 344 (6187):998–1001.

Gladstone, Melissa, Gillian A Lancaster, Eric Umar, Maggie Nyirenda, Edith Kayira, Nynke R van den Broek, and Rosalind L Smyth. 2010. "The Malawi Developmental Assessment Tool (MDAT): the creation, validation, and reliability of a tool to assess child development in rural African settings." PLoS medicine 7 (5):e1000273.

Grantham-McGregor, Sally M, Christine A Powell, Susan P Walker, and John H Himes. 1991. "Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study." The Lancet 338 (8758):1–5.

Haushofer, Johannes and Jeremy Shapiro. 2016. "The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya." The Quarterly Journal of Economics 131 (4):1973–2042.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." American Economic Review 103 (6):2052–2086.

Heckman, James J, Seong Hyeok Moon, Rodrigo Pinto, Peter A Savelyev, and Adam Yavitz. 2010. "The rate of return to the HighScope Perry Preschool Program." Journal of Public Economics 94 (1-2):114–128.

Herth, Kaye. 1992. "Abbreviated instrument to measure hope: Development and psychometric evaluation." Journal of Advanced Nursing 17 (10):1251–1259.

Holla, Alaka, Maria Magdalena Bendini, Lelys Ileana Dinarte Diaz, and Iva Trako. 2021. "Is Investment in Preprimary Education Too Low? Lessons from (Quasi) ExperimentalEvidence across Countries." Tech. rep., The World Bank.

Jeong, Joshua, Dana Charles McCoy, and Aisha K. Yousafzai. 2021. "Cost-Effectiveness and Economic Returns of Group-Based Parenting Interventions to Promote Early Childhood Development: Results from a Randomized Controlled Trial in Rural Kenya." PLOS Medicine 18 (9):e1003746.

Kansikas, Carolina, Anandi Mani, and Paul Niehaus. 2023. "Customized cash transfers: Financial lives and cash-flow preferences in rural Kenya." .

Kes, Aslihan and Hema Swaminathan. 2006. "Gender and time poverty in sub-Saharan Africa." Gender, time use, and poverty in sub-Saharan Africa 13.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental analysis of neighborhood effects." Econometrica 75 (1):83–119.

Kroenke, Kurt, Robert L Spitzer, and Janet BW Williams. 2001. "The PHQ-9: validity of a brief depression severity measure." Journal of general internal medicine 16 (9):606–613.

Leroy, Jef L., Marie Ruel, Jean-Pierre Habicht, and Edward A. Frongillo. 2014. "Linear Growth Deficit Continues to Accumulate Beyond the First 1000 Days in Low- and Middle-Income Countries: Global Evidence from 51 National Surveys." The Journal of Nutrition 144 (9):1460–1466.

Little, Melissa T., Keetie Roelen, Berit C. L. Lange, Janina I. Steinert, Alexa R. Yakubovich, Lucie Cluver, and David K. Humphreys. 2021. "Effectiveness of Cash-Plus Programmes on Early Childhood Outcomes Compared to Cash Transfers Alone: A Systematic Review and Meta-Analysis in Low- and Middle-Income Countries." PLOS Medicine 18 (9):e1003698.

Lybbert, Travis J. and Bruce Wydick. 2018. "Poverty, Aspirations, and the Economics of Hope." Economic Development and Cultural Change 66 (4):709–753.

Macours, Karen, Norbert Schady, and Renos Vakis. 2012. "Cash transfers, behavioral changes, and cognitive development in early childhood: evidence from a randomized experiment." American Economic Journal: Applied Economics 4 (2):247–273.

Manley, James, Harold Alderman, and Ugo Gentilini. 2022. "More evidence on cash transfers and child nutritional outcomes: a systematic review and meta-analysis." BMJ Global Health 7 (4).

Manley, James, Seth Gitter, and Vanya Slavchevska. 2013. "How effective are cash transfers at improving nutritional status?" World Development 48:133–155.

McIntosh, Craig and Andrew Zeitlin. 2024. "Cash versus kind: Benchmarking a child nutrition program against unconditional cash transfers in Rwanda." The Economic Journal 134 (664):3360–3389.

McKenzie, David, Aakash Mohpal, and Dean Yang. 2022. "Aspirations and Financial Decisions: Experimental Evidence from the Philippines." Journal of Development Economics 156:102846.

Meghir, Costas, Orazio Attanasio, Pamela Jervis, Monimalika Day, Prerna Makkar, Jere Behrman, Prachi Gupta, Rashim Pal, Angus Phimister, Nisha Vernekar et al. 2023. "Early stimulation and enhanced preschool: a randomized trial." Pediatrics 151 (Supplement 2):e2023060221H.

Morgan, Kari Lock and Donald B Rubin. 2012. "Rerandomization to improve covariate balance in experiments." The Annals of Statistics 40 (2):1263–1282.

NSO and ICF. 2017. "Malawi demographic and health survey 2015–16." Zomba, Malawi, and Rockville, Maryland, USA. NSO and ICF.," .

Orkin, Kate, Rob Garlick, Mahreen Mahmud, Richard Sedlmayr, Johannes Haushofer, and Stefan Dercon. 2023. "Aspiring to a better future: Can a simple psychological intervention reduce poverty?" Tech. rep., National Bureau of Economic Research.

Özler, Berk, Lia CH Fernald, Patricia Kariger, Christin McConnell, Michelle Neuman, and Eduardo Fraga. 2018. "Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial." Journal of Development Economics 133:448–467.

Pisani, Lauren, Ivelina Borisova, and Amy Jo Dowd. 2018. "Developing and validating the international development and early learning assessment (IDELA)." International Journal of Educational Research 91:1–15.

Poll, Moritz. 2026a. "Micro-enterprise saturation: Critical mass or overcrowding?" Tech. rep.

———. 2026b. "Poverty alleviation at high frequency." Tech. rep.

———. 2026c. "Seasons of change: Experimental evidence on the optimal timing for poverty alleviation." Tech. rep.

Raza, Wameq A., Ellen Van de Poel, and Tom Van Ourti. 2018. "Impact and Spill-Over Effects of an Asset Transfer Program on Child Undernutrition: Evidence from a Randomized Control Trial in Bangladesh." Journal of Health Economics 62:105–120.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, depression, and anxiety: Causal evidence and mechanisms." Science 370 (6522):eaay0214.

Ritzwoller, David M, Joseph P Romano, and Azeem M Shaikh. 2025. "Randomization inference: Theory and applications." Journal of Political Economy: Microeconomics .

Rodriguez-Martinez, Andrea, Bin Zhou, Marisa K Sophiea, James Bentham, Christopher J Paciorek, Maria LC Iurilli, Rodrigo M Carrillo-Larco, James E Bennett, Mariachiara Di Cesare, Cristina Taddei et al. 2020. "Height and body-mass index trajectories of school-aged children and adolescents from 1985 to 2019 in 200 countries and territories: A pooled analysis of 2181 population-based studies with 65 million participants." The Lancet 396 (10261):1511–1524.

Schreiner, Mark. 2019. "Simple Poverty Scorecard® Tool Malawi." .

Suh, Hyun Soo. 2026. "Parental work flexibility and child development." Tech. rep.

Tamim, Sheikh Arman, Narayan Das, Imran Matin, and Munshi Sulaiman. 2026. "Poverty Reduction and Survival: Long-Term Experimental Evidence from Bangladesh." Tech. rep.

Victora, Cesar G., Mercedes de Onis, Pedro C. Hallal, Monika Blössner, and Roger Shrimpton. 2010. "Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions." Pediatrics 125 (3):e473–e480.

Waldman, Marcus, Dana Charles McCoy, Jonathan Seiden, Jorge Cuartas, and Günther Fink. 2021. "The Caregiver Reported Early Development Instruments (CREDI): A psychometric and practical evaluation." International Journal of Behavioral Development 45 (6):562–574.

Walker, Susan P, Susan M Chang, Christine A Powell, and Sally M Grantham-McGregor. 2005. "Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study." The Lancet 366 (9499):1804–1807.

World Bank. 2021. "The State of Economic Inclusion Report." Tech. rep., Partnership for Economic Inclusion.

———. 2024. "The State of Economic Inclusion Report." Tech. rep., Partnership for Economic Inclusion.

Young, Alwyn. 2019. "Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results." The Quarterly Journal of Economics 134 (2):557–598.

Yousafzai, Aisha K, Muneera A Rasheed, Arjumand Rizvi, Robert Armstrong, and Zulfiqar A Bhutta. 2014. "Effect of integrated responsive stimulation and nutrition interventions in the Lady Health Worker programme in Pakistan on child development, growth, and health outcomes: a cluster-randomised factorial effectiveness trial." The Lancet 384 (9950):1282–1293.

Zhou, Jin, James Heckman, Bei Liu, and Mai Lu. 2026. "The Impact of a Prototypical Home Visiting Program on Child Skills." Journal of Labor Economics 44 (1):119–148.

Zhou, Jin, James J Heckman, Bei Liu, Mai Lu, Susan M Chang, and Sally Grantham-McGregor. 2023. "Comparing china REACH and the Jamaica home visiting program." Pediatrics 151 (Supplement 2):e2023060221I.

Appendix

A Descriptives

A.1 Setting

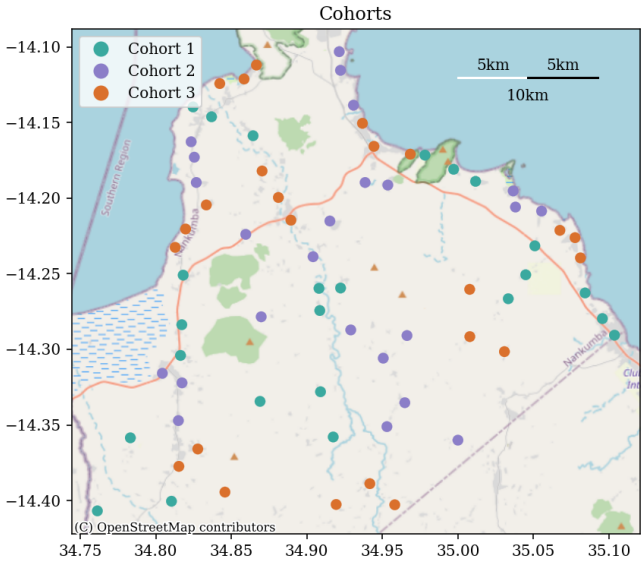


Figure A1: **Clusters** by Cohort

Note: This map shows the 72 study clusters of 26 households each that make up the study sample, color-coded by the cohort they are assigned to.

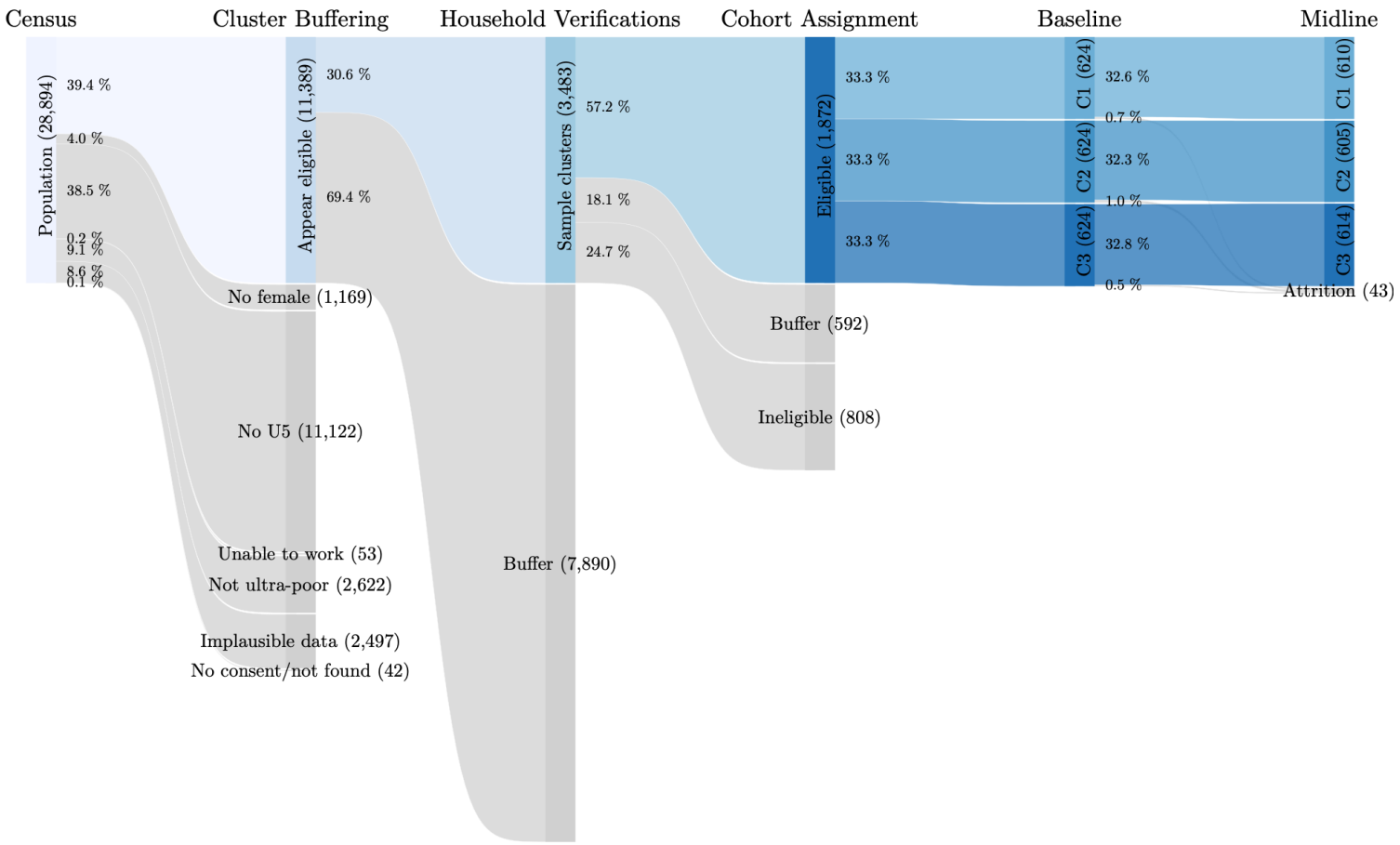


Figure A2: Sampling Funnel

Note: This figure describes how the sample is selected from the census population. Starting from the census population of all 28,894 households, about 60% were deemed ineligible, mostly for lack of a child under age five. Among apparently eligible households, 271 clusters of 40-50 households were formed by proximity using k-means clustering. A second custom algorithm then selected 72 of these clusters in such a way as to maximize the distance between the nearest two clusters, in order to minimize spillover concerns (cf. Appendix Figure A1). The other clusters serve as buffer zones. Among the selected clusters, household verifications were conducted that resulted in some households being deemed ineligible under the same criteria as in the census, while selecting the 26 eligible households closest to the cluster centroid and excluding the remainder as additional buffer households on the periphery of the clusters. The 72 clusters were grouped into triplets by proximity and assigned an implementing staff. They were also assigned at random to three cohorts to undergo the intervention at different times in the agricultural season. The full sample was surveyed at baseline and the graph shows attrition in the subsequent survey rounds.

Table A1: Study Population

	(1) Count
HOUSEHOLDS	
Number of surveys conducted	28,853
Female primary adult present	27,684
U5 present	16,882
No children present	3,254
INDIVIDUALS	
Population	138,548
Adults	65,162
Children	73,386
U5s	20,998
Number of villages	162
Number of CBOs	25
Number of ADCs	5

Census of all households in Mangochi district's subdistrict ("Traditional Authority") Nankumba. The urban center Monkey Bay and the tourist destination Cape Maclear were omitted from this census as well as an area in the South-West of Nankumba where another NGO is implementing a livelihood program. Area Development Committees (ADCs) are administrative subdivisions of a subdistrict, while Community-Based Organizations (CBOs) are associations of several villages organizing development initiatives or providing local public goods together.

A.2 Study implementation

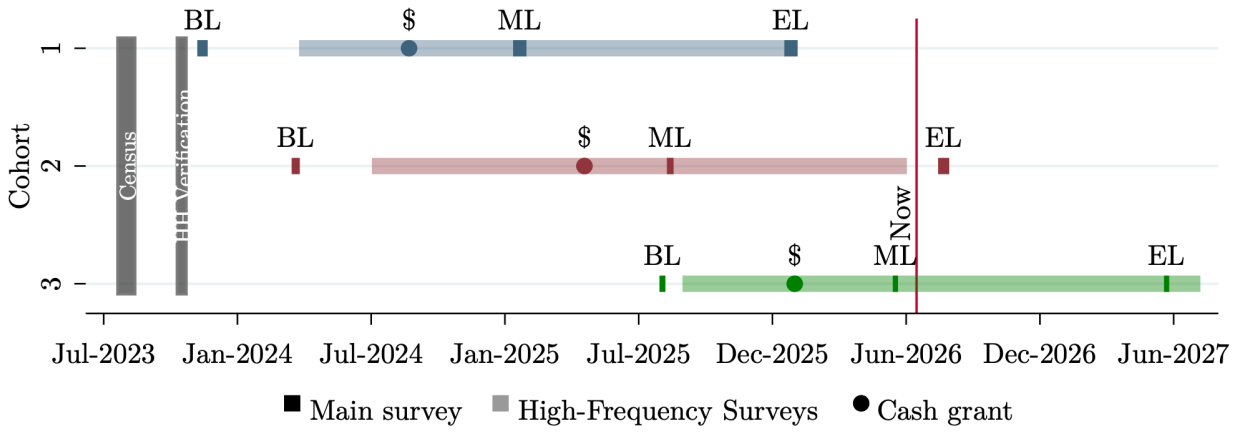


Figure A3: Study Timeline

Note: This figure describes the study timeline. A census of the study area was conducted in August 2023. Household eligibility verifications followed in October 2023. The baseline survey for cohort 1 was conducted in December 2023 and high-frequency surveys in this cohort started in April 2024. The large one-off cash grant was paid out in August 2024. The midline survey was collected in January 2025, followed a year later by the endline survey.

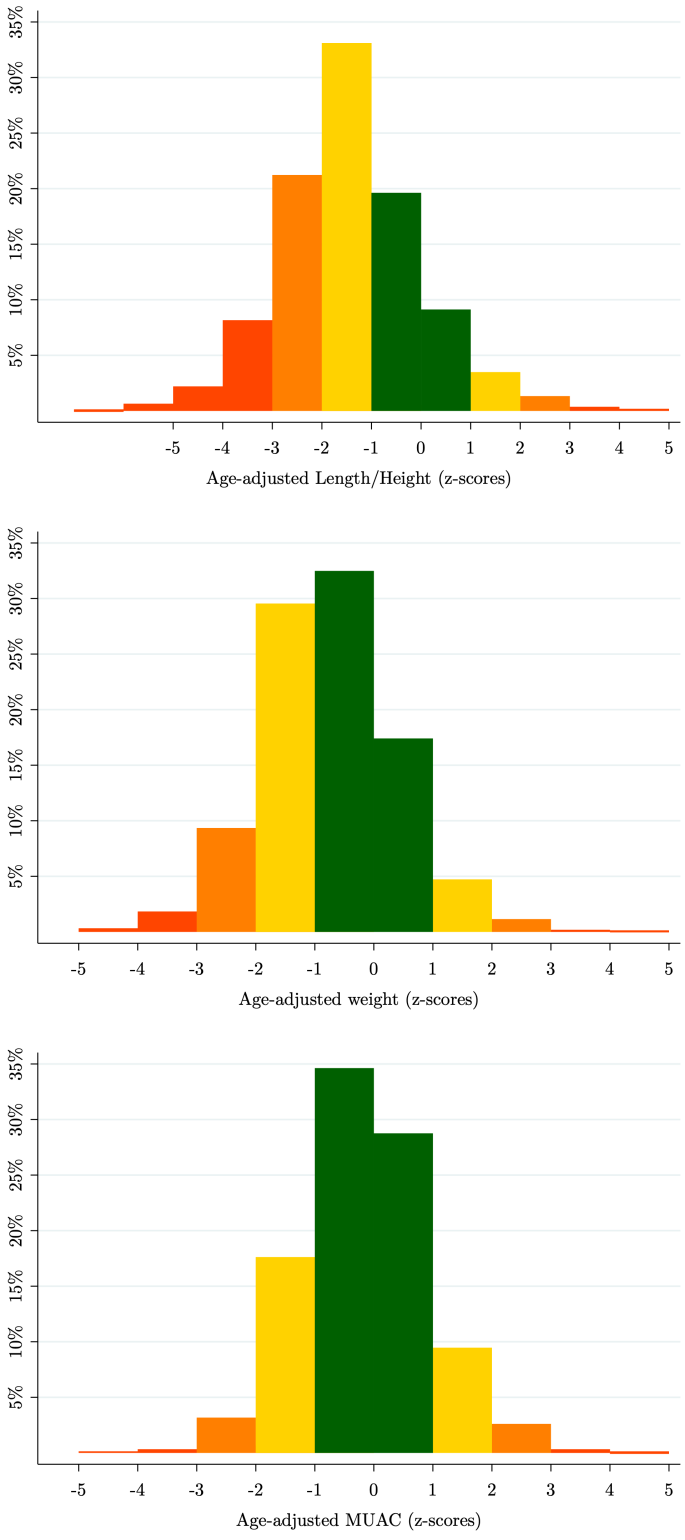


Figure A4: Anthropometrics at Baseline

Note: This figure shows baseline height-for-age (HAZ), weight-for-age (WAZ), and mid-upper arm circumference-for-age z-scores (MAZ) as referenced against WHO lookup tables for all study children born after 01 August 2018, i.e. under age five at enrollment.

Table A2: Sample sizes

	Cohort 1			Cohort 2			Cohort 3		
	Base	Mid	End	Base	Mid	End	Base	Mid	End
Individuals	3318	3260	3088	3338	3303		3403	3393	
BL-U5	855	925	945	874	968		995	1054	
Households	624	610	564	624	605		624	614	
Parenting	624	610	564	624	603		624	610	
U5	814	673	568	765	635		662	606	
MUAC	782	609	490	731	576		632	564	
Height	797	623	506	743	593		641	578	
Weight	793	626	513	737	598		652	578	
HH w/ 0-3	406	335	265	409	291		296	244	
CREDI			245		286		257	234	
HH w/ 2-9	553	577	559	574	603		618	613	
IDELA		496	344		382		405	430	
U6	932	855	724	909	805		815	759	
MDAT			624						

Notes: This table shows the sample sizes for the various survey waves and assessments. Households indicate the number of surveyed households and Individuals the number of individuals comprised in them. BL-U5 is the number of children under age 5 upon enrollment (eligibility criterion). U5 is the number of children under age 5 during the survey (theoretical maximum sample size for anthropometrics). This includes new births. HH w/ 0-3 is the number of households with a child in the age range 0-3 (maximum theoretical sample size for CREDI, one child per household assessed). HH w/ 2-9 is the number of households with a child in the age range 2-9 (maximum theoretical sample size for IDELA, one child per household assessed). The remainder of rows are realized sample sizes for the various assessments.

Table A3: Child Outcomes

	Cohort 1			Cohort 2			Cohort 3		
	Base	Mid	End	Base	Mid	End	Base	Mid	End
MUAC	✓	✓	✓	✓	✓	✓	✓	✓	✓
Height	✓	✓	✓	✓	✓	✓	✓	✓	✓
Weight	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parenting	✓	✓	✓	✓	✓	✓	✓	✓	✓
IDELA		✓	✓		✓	✓	✓	✓	✓
CREDI			✓		✓	✓	✓	✓	✓
MDAT			✓			✓			✓

Notes: ✓ indicates that the outcome is measured for the corresponding cohort and survey wave. Health outcomes include caregiver-reported morbidity and anthropometric measures (height, weight, and mid–upper arm circumference). CREDI is a caregiver-reported developmental assessment, while IDELA and MDAT are directly child-assessed. MUAC is additionally assessed approximately monthly in high-frequency surveys (HFS).

A.3 IDELA z-scoring

While the CREDI and MDAT assessments have international benchmark data comparable to WHO lookup tables for age-adjusted anthropometrics (operationalized through online applications), the IDELA scale does not provide such a benchmark. We therefore compute it ourselves. In doing so, it is important to estimate not just the mean but also the standard deviation by age since the raw data tend to fan out with age and for some tasks fan back in for the oldest children. Figure A5 describes the process for the overall (top panel), cognitive subdomain (middle panel), and non-cognitive subdomain (bottom panel) score. In a first step, we create non-parametric, binned estimates of both mean and standard deviation. Bins have a width of one year with a six-month step size. The oldest age bin is two years wide due to small sample size. The left panel shows these binned estimates as well as a smoothed logistic function that is used for z-scoring in the next step. A logistic function has the advantage of being naturally bounded above and below (like the test scores) but allowing the data to speak to where the bounds lie and at what age and how gradually they should be reached. The middle panel shows how the resulting smoothed location and scale parameters fit the raw data. The right panel shows the resulting z-scores.

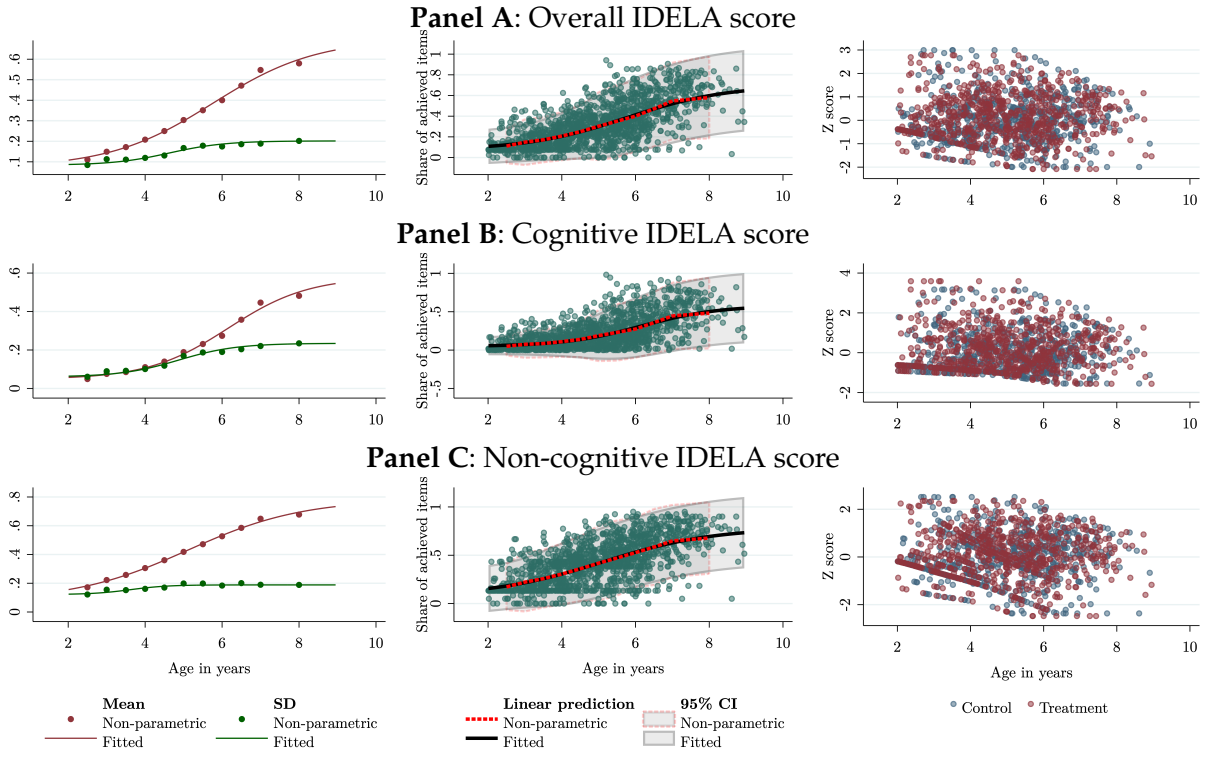


Figure A5: IDELA z-scoring (midline)

Note: This figure summarizes z-scoring of the IDELA assessment scores for children aged 2-9. The left shows estimated age-adjusted mean and standard deviations based on binned non-parametric averages. The middle shows how the raw data fall relative to the fitted scores. The right shows the resulting z-scores.

A.4 High-Frequency Time Use Questions

- In the past seven days, how many hours have you spent on Yamba Malawi (implementing partner) program activities?
- In the past seven days, how many hours have you spent on household chores (including cleaning, washing, meal preparation, shopping, etc. but not childcare)?
- In the past seven days, how many hours have you spent on traveling or transport (include all forms of transport including walking)?
- In the past seven days, how many hours have you spent on childcare?
- In the past seven days, how many hours have you spent on leisure or resting by yourself (but not sleeping)?
- In the past seven days, how many hours have you spent on social, community, or religious activities?
- In the past seven days, how many hours have you spent idle (not resting or sleeping)?
- Over the past seven days, how many hours did you (Program participant) spend on this income source?

B Factor Loadings of Parental Investments

Table B1 reports the confirmatory factor analysis (CFA) loadings for the parental investment measurement system at baseline and midline, respectively. The CFA is estimated on the control group only (Section 3). All loadings are relative to the normalizing item (clothing for material investment, reading to the child for parenting quality), which is fixed at 1.

Within the material investment factor, shoes and meals load most strongly. The parenting quality factor is dominated by educational stimulation items: counting, letters, colors, stories, and religious teaching all load above 0.8 at both waves. Parenting practice items—affection, discipline, supervision, and encouragement—load weakly (below 0.3), indicating that these behaviors share relatively little common variation with the educational stimulation dimension. The factor correlation is moderate and stable across waves, confirming that the two investment dimensions are related but distinct.

Table B1: CFA Factor Loadings

Item	**Baseline**		**Midline**	
	Material investment	Parental quality	Material investment	Parental quality
All children own a pair of shoes	1.000	—	1.000	—
All children own a blanket	0.520	—	0.724	—
Household owns toys	0.639	—	0.637	—
All children own 3+ sets of clothes	0.217	—	0.138	—
All children 3+ meals	0.764	—	0.751	—
Reads to child	—	1.000	—	1.000
Teaches child counting/numbers	—	1.099	—	1.043
Teaches letters/sounds	—	1.146	—	1.068
Teaches shapes	—	1.131	—	1.034
Teaches hygiene	—	0.734	—	0.418
Teaches religion to child	—	0.745	—	0.596
Tells child stories	—	0.781	—	0.674
Sings songs to child	—	0.405	—	0.496
Takes child on outings	—	0.163	—	0.491
Plays with child	—	0.102	—	0.256
Let child partake in chores	—	0.654	—	0.525
Showed affection	—	0.022	—	0.097
Pleased with child	—	0.253	—	0.099
Does not spank child	—	-0.198	—	-0.021
Constructive discipline	—	0.195	—	0.178
Gently encouraged eating	—	0.196	—	0.146
Child not left unsupervised	—	-0.055	—	0.123
Factor correlation	0.391		0.308	
RMSEA	0.053		0.059	
N	936		900	

Notes: This table shows the confirmatory factor loadings for the material investment and parental quality factors for all survey waves.

C Additional Results

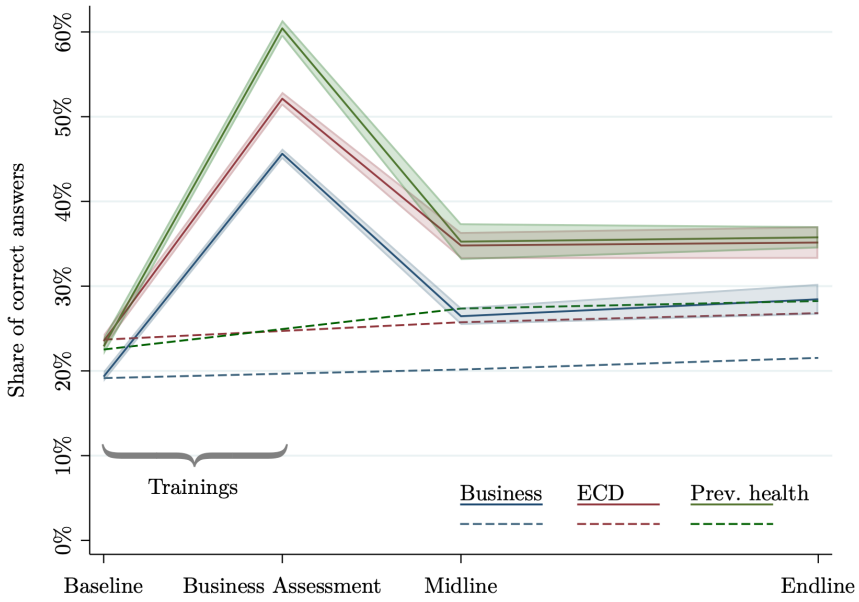


Figure C1: Knowledge gain and fade-out

Note: This figure shows the gain and subsequent fade-out in knowledge from the various training topics. Program participants are asked open-ended questions on each lesson topic ("How do you prevent coughs and other respiratory infections?") while enumerators check off correct answers ("Ventilation", "Handwashing", "Covering mouth and nose when coughing/sneezing", ...). Both treatment (solid line) and control group (dashed line) are assessed at every main-wave survey but treatment households are additionally assessed right at the end of their training. Knowledge gains at that point are remarkably high but fade out quickly and stabilize at a level that is modestly but significantly higher than the control group.

Table C1: Baseline Sample Balance

	(1) Control Mean (SD)	(2) Treatment (SE)
HOUSEHOLD AND PARTICIPANT CHARACTERISTICS		
Household size	5.432	-0.118
	(1.827)	(0.097)
Age	31.6	-0.564
	(8.443)	(0.424)
Years of education	5.657	-0.122
	(2.880)	(0.191)
Literate	0.692	0.005
	(0.462)	(0.027)
LIVELIHOOD AND ASSETS		
Hours worked (7 days)	22	-0.364
	(17.9)	(0.783)
HH income (12 months, MWK)	475,629	11,315
	(480,571)	(47,515)
Total savings, assets, livestock, minus debt (MWK)	256,874	-17,471
	(365,607)	(24,353)
Largest purchase (12 months, MWK)	38,607	-2,816
	(121,528)	(6,344)
POVERTY AND FOOD SECURITY		
Thatch / tarp / no roof	0.746	0.021
	(0.436)	(0.023)
Food insecurity score (0-8)	7.532	0.049
	(1.341)	(0.064)
HEALTH AND NON-HEALTH SHOCKS		
Health episodes (12 months)	15.6	-0.578
	(12.8)	(0.920)
Non-health shocks (12 months)	20.6	-0.377
	(18.1)	(0.853)
NOT PRE-SPECIFIED		
Total number of children	3.298	-0.092
	(1.551)	(0.086)
Pregnancies per household	0.091	0.003
	(0.291)	(0.013)
Female empowerment (0-1)	0.472	0.011
	(0.268)	(0.017)
Height-for-age (HAZ)	-1.423	0.017
	(1.403)	(0.071)
Weight-for-age (WAZ)	-0.761	-0.071
	(1.096)	(0.065)
MUAC-for-age (MAZ)	-0.171	-0.140**
	(1.069)	(0.062)
CREDI score (z)	-0.974	-0.196
	(1.461)	(0.165)
Child attends pre-school	0.412	-0.050
	(0.492)	(0.043)
Material investment (SD)	0.000	-0.032
	(1.000)	(0.069)
Parenting quality (SD)	0.000	-0.024
	(1.000)	(0.063)
Observations		1,872
F		0.58
$p(F = 0)$		0.851

This table describes the baseline balance. The main panel reports the control group means and coefficients of the treatment indicator in a regression predicting each baseline outcome variable. At the bottom of the table, an F statistic is reported for a regression in which all outcomes jointly predict the treatment. Only the pre-specified variables are included in the F-test. Standard errors are clustered at the household cluster level and reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: Impacts on Material Investment (Detail)

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
All children own 3+ sets of clothes	0.975	0.020***	1829
	(0.158)	(0.006)	✓
		[0.001]	
Household owns toys	0.282	0.336***	1828
	(0.450)	(0.033)	✓
		[0.000]	
All children own a blanket	0.231	0.362***	1829
	(0.422)	(0.030)	✓
		[0.000]	
All children own a pair of shoes	0.555	0.252***	1829
	(0.497)	(0.021)	✓
		[0.000]	
All children 3+ meals	0.363	0.407***	1829
	(0.481)	(0.023)	✓
		[0.000]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: Impacts on Child-Related Expenditure

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Child health spending (MWK)	445	237***	1530
	(823)	(76.9)	×
		[0.003]	
Child food spending (MWK)	12,327	5,180***	1530
	(12,188)	(1,071)	×
		[0.000]	
Child clothes/toys spending (MWK)	1,162	2,630***	1530
	(1,971)	(306)	×
		[0.000]	
Child education spending (MWK)	1,308	932***	1530
	(2,224)	(180)	×
		[0.000]	
Child-directed spending (MWK/month)	15,243	8,978***	1530
	(14,489)	(1,338)	×
		[0.000]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C4: Impacts on Parenting Quality (Self-Reported, Detail)

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Parenting quality (SD)	0.000	0.415***	1823
	(1.000)	(0.049)	✓
		[0.000]	
Reads to child	0.522	0.174***	1828
	(0.500)	(0.021)	✓
		[0.000]	
Teaches child counting/numbers	0.659	0.130***	1828
	(0.474)	(0.021)	✓
		[0.000]	
Teaches letters/sounds	0.552	0.146***	1828
	(0.498)	(0.023)	✓
		[0.000]	
Teaches shapes	0.534	0.154***	1828
	(0.499)	(0.023)	✓
		[0.000]	
Teaches hygiene	0.889	0.044***	1828
	(0.314)	(0.016)	✓
		[0.009]	
Teaches religion to child	0.819	0.076***	1823
	(0.385)	(0.016)	✓
		[0.000]	
Tells child stories	0.758	0.097***	1828
	(0.428)	(0.022)	✓
		[0.000]	
Sings songs to child	0.861	0.089***	1828
	(0.346)	(0.017)	✓
		[0.000]	
Takes child on outings	0.806	0.069***	1828
	(0.396)	(0.020)	✓
		[0.001]	
Plays with child	0.930	0.037***	1828
	(0.255)	(0.010)	✓
		[0.001]	
Let child partake in chores	0.590	0.078***	1828
	(0.492)	(0.022)	✓
		[0.001]	
Showed affection	0.955	0.020**	1828
	(0.208)	(0.008)	✓
		[0.025]	
Pleased with child	0.909	0.039***	1828
	(0.288)	(0.010)	✓
		[0.001]	
Does not spank child	0.647	0.100***	1828
	(0.478)	(0.024)	✓
		[0.000]	
Constructive discipline	0.462	0.115***	1828
	(0.499)	(0.023)	✓
		[0.000]	
Gently encouraged eating	0.732	0.056**	1828
	(0.443)	(0.020)	✓
		[0.009]	
Child not left unsupervised	0.711	0.052***	1828
	(0.454)	(0.017)	✓
		[0.005]	
Knows child's weight	0.696	0.054***	1828
	(0.460)	(0.019)	✓
		[0.007]	
Took child to check-up (6 months)	0.229	0.010	1828
	(0.421)	(0.016)	✓
		[0.525]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C5: Impacts on Parenting Quality (Observations, Detail)

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Parenting quality observed (SD)	0.000	0.277***	873
	(1.000)	(0.050)	(✓)
		[0.000]	
Keeps child busy	0.813	0.052**	873
	(0.390)	(0.025)	(✓)
		[0.041]	
Teaches child words	0.467	−0.018	873
	(0.499)	(0.027)	(✓)
		[0.505]	
Shows affection	0.931	0.013	873
	(0.254)	(0.019)	(✓)
		[0.488]	
Spanks child	0.053	−0.004	873
	(0.224)	(0.013)	(✓)
		[0.774]	
Shouts at child	0.146	−0.019	873
	(0.353)	(0.023)	(✓)
		[0.394]	
Positive emotions toward child	0.830	0.051*	873
	(0.376)	(0.026)	(✓)
		[0.052]	
Responds to sounds	0.967	0.005	873
	(0.180)	(0.013)	(✓)
		[0.672]	
Aware of child needs	0.646	0.057*	873
	(0.479)	(0.030)	(✓)
		[0.073]	
Keeps child in view	0.945	0.007	873
	(0.228)	(0.014)	(✓)
		[0.637]	
Child looks clean	0.541	0.230***	873
	(0.499)	(0.030)	(✓)
		[0.000]	
Home is clean	0.758	0.138***	873
	(0.429)	(0.021)	(✓)
		[0.000]	
Inside safe to play	0.682	0.121***	873
	(0.466)	(0.025)	(✓)
		[0.000]	
Outside safe to play	0.648	0.133***	873
	(0.478)	(0.028)	(✓)
		[0.000]	
Child uses toy	0.196	0.240***	873
	(0.398)	(0.032)	(✓)
		[0.000]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C6: Absence of knowledge spillovers

	(1) Control mean (SD)	(2) Treatment Effect	(3) Spillover Control	(4) N (BL)
Business knowledge (0-1)	0.202	0.065***	0.004	1829
	(0.074)	(0.006)	(0.007)	✓
		[0.000]	[0.552]	
ECD knowledge (0-1)	0.257	0.095***	0.012	1829
	(0.099)	(0.009)	(0.010)	✓
		[0.000]	[0.239]	
Preventive health knowledge (0-1)	0.274	0.080***	0.005	1829
	(0.137)	(0.013)	(0.012)	✓
		[0.000]	[0.685]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C7: Women's Time Allocation (before and after cash grant)

	(1) Control mean (SD) **Pre-grant**	(2) Control mean (SD) **Post-grant**	(3) Treatment Effect **Pre-grant**	(4) Treatment Effect **Post-grant**	(5) Difference p-value	(6) N (BL)
Program involvement	1.544	1.251	4.851***	2.369***	[0.000]***	3064
	(2.324)	(1.407)	(0.542)	(0.487)		×
			[0.000]	[0.000]		
Business	7.006	7.581	4.070***	7.824***	[0.001]***	3064
	(8.887)	(8.410)	(0.827)	(0.741)		×
			[0.000]	[0.000]		
Farming	1.211	2.818	-0.048	1.337	[0.193]	3064
	(3.284)	(5.176)	(0.280)	(1.014)		×
			[0.875]	[0.194]		
Ganyu	9.411	6.561	-3.624***	-4.984***	[0.102]	3064
	(9.375)	(7.796)	(0.684)	(0.922)		×
			[0.000]	[0.000]		
Household chores	22.9	21.8	-0.407*	-0.645***	[0.458]	3064
	(7.288)	(7.112)	(0.209)	(0.213)		×
			[0.062]	[0.004]		
Childcare	21.8	22.2	-0.630***	-0.504*	[0.700]	3064
	(8.490)	(8.881)	(0.185)	(0.249)		×
			[0.001]	[0.053]		
Travel	10.5	10.5	-0.781***	-0.430***	[0.111]	3064
	(5.264)	(4.911)	(0.191)	(0.115)		×
			[0.001]	[0.001]		
Leisure	8.756	8.708	-0.869***	-0.729***	[0.640]	3064
	(4.297)	(3.999)	(0.239)	(0.242)		×
			[0.000]	[0.002]		
Social	5.983	6.044	-0.072	-0.452*	[0.065]*	3064
	(3.891)	(3.241)	(0.121)	(0.227)		×
			[0.579]	[0.048]		
Idle	6.102	6.599	-0.546**	-0.718***	[0.489]	3064
	(4.298)	(4.100)	(0.240)	(0.198)		×
			[0.028]	[0.001]		
Sleep & Meals	72.6	73.8	-2.043***	-3.031***	[0.021]**	3064
	(6.877)	(5.218)	(0.405)	(0.477)		×
			[0.000]	[0.000]		

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C8: Childcare Arrangements While Working

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Left the baby home alone	0.026	−0.018	1528
	(0.124)	(0.016)	×
		[0.123]	
~ with another child younger than 10 years old	0.023	−0.017***	1528
	(0.077)	(0.005)	×
		[0.000]	
~ with another child that is 10 years or older	0.085	−0.035***	1528
	(0.156)	(0.011)	×
		[0.001]	
~ with the father	0.027	−0.009	1528
	(0.095)	(0.015)	×
		[0.829]	
~ with another adult family member	0.300	−0.074***	1528
	(0.293)	(0.023)	×
		[0.001]	
~ with an adult neighbor	0.071	−0.019**	1528
	(0.136)	(0.007)	×
		[0.007]	
Took the baby with you	0.268	−0.065**	1528
	(0.305)	(0.031)	×
		[0.037]	
Brought baby to the CBCC or another childcare facility	0.199	0.238***	1528
	(0.286)	(0.044)	×
		[0.000]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C9: Impacts on Mothers' Mental Health

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Mother's PHQ-9 depression score (0-27)	4.780	−2.808***	614
	(3.419)	(0.325)	×
		[0.000]	
Little interest or pleasure in doing things	0.711	−0.437***	614
	(0.581)	(0.052)	×
		[0.000]	
Feeling down, depressed, or hopeless	0.813	−0.520***	614
	(0.629)	(0.053)	×
		[0.000]	
Trouble falling or staying asleep, or sleeping too much	0.633	−0.364***	614
	(0.636)	(0.052)	×
		[0.001]	
Feeling tired or having little energy	0.630	−0.251***	614
	(0.554)	(0.046)	×
		[0.003]	
Poor appetite or overeating	0.521	−0.298***	614
	(0.574)	(0.052)	×
		[0.002]	
Feeling like a failure	0.489	−0.331***	614
	(0.602)	(0.039)	×
		[0.000]	
Trouble concentrating on things	0.416	−0.280***	614
	(0.544)	(0.046)	×
		[0.000]	
Moving slowly or acting fidgety	0.351	−0.182***	614
	(0.572)	(0.048)	×
		[0.008]	
Suicidal ideation	0.216	−0.145***	614
	(0.472)	(0.027)	×
		[0.002]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C10: Impacts on Mothers' Hope

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Herth Hope Index (1-5)	3.896	0.207***	1828
	(0.533)	(0.041)	(✓)
		[0.000]	
I have a positive outlook toward life	4.035	0.337***	1828
	(1.032)	(0.062)	(✓)
		[0.000]	
I have short and long range goals	4.081	0.270***	1828
	(0.909)	(0.054)	(✓)
		[0.000]	
I feel all alone	3.029	−0.064	1828
	(1.391)	(0.072)	(✓)
		[0.387]	
I can see possibilities in the midst of difficulties	3.878	0.244***	1828
	(1.033)	(0.061)	(✓)
		[0.000]	
I have a faith that gives me comfort	4.058	0.204***	1828
	(0.935)	(0.062)	(✓)
		[0.003]	
I feel scared about my future	3.132	−0.367***	1828
	(1.377)	(0.080)	(✓)
		[0.000]	
I can recall happy/joyful times	4.000	0.169***	1828
	(0.971)	(0.058)	(✓)
		[0.008]	
I have deep inner strength	3.961	0.197***	1828
	(0.935)	(0.053)	(✓)
		[0.001]	
I am able to give and receive caring/love	4.259	0.136**	1828
	(0.749)	(0.055)	(✓)
		[0.017]	
I have a sense of direction	4.129	0.175***	1828
	(0.872)	(0.058)	(✓)
		[0.005]	
I believe that each day has potential	4.224	0.149***	1828
	(0.801)	(0.049)	(✓)
		[0.007]	
I feel my life has value and worth	4.289	0.144***	1828
	(0.831)	(0.048)	(✓)
		[0.007]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C11: Impacts on Female Empowerment

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Female empowerment (0-1)	0.533	0.009	1367
	(0.246)	(0.013)	(✓)
		[0.490]	
How much of the household's income to save?	−0.476	0.130*	1367
	(1.673)	(0.074)	(✓)
		[0.087]	
Meal portions when food reserves are low?	−0.047	0.057	1367
	(1.736)	(0.084)	(✓)
		[0.514]	
What to do if a child falls sick?	0.364	0.087	1367
	(1.667)	(0.083)	(✓)
		[0.305]	
Any decisions about children's schooling?	−0.035	0.189*	1367
	(1.629)	(0.103)	(✓)
		[0.078]	
How should children be disciplined?	0.234	−0.012	1367
	(1.574)	(0.104)	(✓)
		[0.913]	
Whether or not to use a method to avoid having children?	0.898	−0.113	1367
	(1.452)	(0.083)	(✓)
		[0.185]	
Whether to have another child?	−0.015	−0.054	1367
	(1.666)	(0.107)	(✓)
		[0.629]	
Domestic violence justifiable	0.446	−0.028	1785
	(0.497)	(0.022)	(✓)
		[0.186]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C12: Impacts on Fertility

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Total number of children	3.360	−0.005	1829
	(1.461)	(0.040)	✓
		[0.881]	
Total number of children under five	1.605	0.002	1829
	(0.679)	(0.028)	✓
		[0.948]	
Births	0.137	−0.002	1829
	(0.554)	(0.034)	✓
		[0.949]	
Program participant pregnant	0.059	0.019	1828
	(0.235)	(0.012)	✓
		[0.123]	
Pregnancies per household	0.074	0.015	1829
	(0.262)	(0.013)	✓
		[0.258]	
Effective contraception	0.683	0.008	1823
	(0.466)	(0.019)	(✓)
		[0.659]	
Ideal number of children (mother)	4.516	−0.021	1828
	(2.144)	(0.088)	(✓)
		[0.808]	
Ideal number of children (father)	4.789	0.064	877
	(1.982)	(0.158)	(✓)
		[0.653]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 10,000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.