INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

RESEARCH PROGRAM ON Policies, Institutions, and Markets
CGIAR
Led by IFPRI

IFPRI Discussion Paper 02090

December 2021

Post-program Impacts of Transfer Programs on Child Development

Experimental evidence from Bangladesh

Akhter Ahmed

Jena Hamadani

Md. Zahidul Hassan

Melissa Hidrobo

John Hoddinott

Bastien Koch

Kalyani Raghunathan

Shalini Roy

Poverty, Health, and Nutrition Division

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Akhter Ahmed (a.ahmed@cgiar.org) is a Senior Research Fellow in the Poverty, Health, and Nutrition Division of the International Food Policy Research Institute (IFPRI) and IFPRI's Country Representative in Dhaka, Bangladesh.

Jena Hamadani (jena@icddrb.org) is an Emeritus Scientist in the Maternal and Child Health Division at icddr,b in Dhaka, Bangladesh.

Md. Zahidul Hassan (z.hassan@databd.org) is the Managing Director at Data Analysis and Technical Assistance (DATA) Limited in Dhaka, Bangladesh.

Melissa Hidrobo (m.hidrobo@cgiar.org) is a Senior Research Fellow in IFPRI's Poverty, Health, and Nutrition Division, Dakar, Senegal.

John Hoddinott (jfh246@cornell.edu) is the H.E. Babcock Professor of Food and Nutrition Economics and Policy in the Division of Nutrition Sciences at Cornell University and a Non-Resident Fellow at IFPRI.

Bastien Koch (b.koch@cgiar.org) is a Research Analyst in IFPRI's Poverty, Health, and Nutrition Division, Washington, DC.

Kalyani Raghunathan (k.raghunathan@cgiar.org) is a Research Fellow in IFPRI's Poverty, Health, and Nutrition Division, New Delhi, India.

Shalini Roy* (s.roy@cgiar.org) is a Senior Research Fellow in IFPRI's Poverty, Health, and Nutrition Division, Washington, DC.

**Corresponding author: s.roy@cgiar.org. 202-862-4640*

Notices

[1] IFPRI Discussion Papers contain preliminary material and research results and are circulated in order to stimulate discussion and critical comment. They have not been subject to a formal external review via IFPRI's Publications Review Committee. Any opinions stated herein are those of the author(s) and are not necessarily representative of or endorsed by IFPRI.

[2] The boundaries and names shown and the designations used on the map(s) herein do not imply official endorsement or acceptance by the International Food Policy Research Institute (IFPRI) or its partners and contributors.

[3] Copyright remains with the authors. The authors are free to proceed, without further IFPRI permission, to publish this paper, or any revised version of it, in outlets such as journals, books, and other publications.

Abstract

Evidence shows transfer programs can improve early childhood development (ECD). However, knowledge gaps remain on how short-term impacts on ECD evolve as children grow older, how program design features and context affect child development impacts over time, and through what pathways such impacts occur. We study the Transfer Modality Research Initiative (TMRI), a 2-year randomized controlled trial in two regions of Bangladesh that provided cash or food transfers, with or without complementary nutrition programming, to mothers of children aged 0-2 years at baseline. Drawing on data collected at 6 months post-program (when children were about 2-4 years old) and at 4 years post-program (when children were about 6-8 years old), we assess post-program impacts of TMRI on children's home environment and development. We find strong post-program impacts on the home environment from cash transfers in the Northern region, particularly when combined with complementary programming, however limited post-program effects on child development outcomes. Improvements found in child development tend to be concentrated in boys. We find few post-program improvements in home environment or child development from food transfers in the Southern region, with or without complementary programming.

Acknowledgments

This work was undertaken as part of the CGIAR Research Program on Policies, Institutions, and Markets (PIM) led by the International Food Policy Research Institute (IFPRI). Funding support for this study was provided by an anonymous donor, the World Food Programme, Cornell University, and the CGIAR Research Program on Policies, Institutions, and Markets. The study builds on research funded by the German Ministry for Economic Cooperation and Development (BMZ), the UK's Department for International Development (DFID), the Swiss Agency for Development and Cooperation (SDC), the United Nations Development Programme (UNDP), and the United States Agency for International Development (USAID). The study was registered in the AEA RCT registry (AEARCTR-0002977) and with ClinicalTrials.gov (ID: NCT03810300), received ethical approval from the Institutional Review Board of the International Food Policy Research Institute (IFPRI), and was reviewed by the Ministry of Food and Disaster Management in Bangladesh. We thank DATA for careful data collection, Wahid Quabili for excellent research assistance, seminar participants for insightful comments, and survey respondents for their participation in the study. All errors are our own.

1. Introduction

Many cash or food transfer programs have dual objectives of addressing the immediate needs of households living in poverty while also breaking the intergenerational transmission of poverty by promoting children's human capital accumulation. A large literature shows these programs' effectiveness in improving short-term outcomes related to poverty reduction, such as food security and dietary diversity (Hidrobo et al. 2018; Bastagli et al. 2016). A substantial literature also demonstrates that cash or food transfer programs can increase human capital accumulation among children in poor households. Much of the evidence on human capital looks at program impacts on school participation, focusing on children aged six years or older (Baird et al. 2014; Fiszbein et al. 2009; Beegle, Coudouel, and Monsalve 2018; Hahn et al. 2018; Ahmed & del Ninno 2002).

A more recent and growing literature shows that transfer programs can also increase pre-school-age children's human capital accumulation (for reviews see Engle et al. 2011, de Walque et al. 2017), and Arriagada et al. 2018). Effects of transfer programs on early childhood development (ECD) can occur through a number of pathways, including via increased income for parental investments in child nutrition, health, and the home environment, as well as reductions in poverty-related parental stress and depression.[1] Combining transfer programs with complementary activities may augment impacts if the complementary programming shifts preferences towards child inputs, increases knowledge of parenting practices around stimulation and improves parental mental health (Attanasio, Cattan, and Meghir 2021; Arriagada et al. 2018). Findings on transfer programs' effects on ECD are encouraging, given evidence that ECD is critical for long-term outcomes (Grantham-McGregor et al. 2007), growing up in poverty is linked to early developmental deficits (Walker et al. 2007; Hamadani et al. 2014), and transfer programs reach large numbers of poor households in many low- and middle-income settings (Ivashchenko et al. 2018).

However, knowledge gaps remain around the effects of cash and food transfer programs on ECD. First, we know little about how short-term program impacts on ECD (or lack thereof) evolve as children grow older. Second, there is limited understanding of how child development impacts at different ages are

[1] The literature presents a nuanced picture on the potential of transfer programs for young children's nutritional status. Evidence suggests that cash or food transfers alone tend to cause limited improvements in anthropometry except in settings with severe deprivation, but that the combination of cash transfers and high-quality complementary nutrition training can significantly improve child anthropometry (Manley et al. 2020; Carneiro et al. 2021; Ahmed et al. 2019).

shaped by program design features – for example, what role is played by complementary programming provided to parents. Third, there is limited evidence on what pathways play a role in transfer programs' impacts on ECD and subsequent human capital indicators – in particular, to what extent effects on home environment and maternal psychological well-being may be important. Lastly, there is scarce evidence on the impacts of transfer programs on child development outside Latin America, however context could matter; for example, in South Asia where son preference has historically prevailed, impacts could plausibly differ by children's sex.

In this paper, we contribute to filling these knowledge gaps using data from a pilot program called the Transfer Modality Research Initiative (TMRI). Implemented in two regions of Bangladesh from 2012-2014, TMRI provided mothers of young children in poor rural households with either monthly cash or food transfers, with or without nutrition behavior change communication (BCC), following a randomized controlled trial (RCT) design. Targeted children (referred to as "index children") were aged 0 to 2 years at the start of the two-year intervention. Thus, all index children received program benefits while under the age of 3 years – an age range that Black et al. (2017) argue particular attention should be given to, given the demonstrated importance of early brain development during this period, the relative lack of attention to ECD at these ages compared to 3 to 5 years, and the potential for various large-scale sectors including social protection to reach children at this life stage. Moreover, the complementary programming aimed to improve children's nutrition in the "first thousand days," also critical for children's neurodevelopment and mental health, as inadequate intake of key nutrients during this period can result in irreversible deficits in brain function (Schwarzenberg et al. 2018).

Taking advantage of the program's RCT design, we assess post-program impacts of TMRI's different treatment arms on home environment and child development. We first assess impacts on index children's home environment and ECD outcomes 6 months after the end of intervention, when these children were aged 2 to 4 years. We then assess impacts on the home environment and age-relevant cognitive, reading, math, and behavioral indicators 4 years after the end of intervention, when the same index children were ages 6 to 8 years. (Henceforth, child outcomes for this age group are referred to simply as child development (CD), as this age range is no longer considered "early childhood.") Our analysis focuses on comparing impacts of the transfers with and without complementary programming, as well as by sex of the child. We find strong impacts of cash transfers in the Northern region on the home environment at 6 months and 4 years post-program, particularly when combined with complementary programming, but limited post-program effects on child development outcomes. Improvements in child development

appear over time at 4yPP and tend to be concentrated in boys. We find few post-program improvements in home environment or child development from food transfers in the Southern region at either 6 months or 4 years post-program, with or without complementary programming.

Our findings relate to results from an RCT in Niger that for approximately 2 years provided either cash transfers alone or cash transfers combined with BCC focused on ECD – including parenting training activities to encourage health, nutrition, psycho-social stimulation, and child protection practices (Premand and Barry 2020). At followup just after intervention, the Niger study found few impacts on child outcomes from cash alone, but large impacts from the cash with BCC for children under 5 years, including on a range of parental practices related to nutrition, health, and stimulation, as well as moderate gains on behavioral indicators. However, it found no significant impacts on anthropometry or cognitive development. Our study on TMRI differs in that the BCC focused specifically on early childhood nutrition, and the ECD analysis focuses on post-program impacts. Prior work on TMRI shows that, just after intervention, all treatment arms improved diets of index children, although adding BCC to transfers led to significantly larger increases in dietary diversity, and the combination of cash and BCC in the North led to significant improvements in anthropometry (Ahmed et al. 2019). Our post-program analysis in this paper suggests that the impacts on diets and nutrition during the program in addition to sustained impacts in the home environment led to some post-program impacts from Cash and Cash+BCC on children's cognitive development and behavior, but impacts took time to emerge and were concentrated among boys.

The paper proceeds as follows. Section 2 introduces the TMRI intervention and study design. Section 3 describes the data and estimation sample for our analysis. Section 4 summarizes our outcomes, while Section 5 describes our empirical approach, and Section 6 presents our main results. Section 7 provides a brief discussion of the results and concludes.

2. Program details and study design

The Transfer Modality Research Initiative (TMRI) conducted two cluster-randomized controlled trials in the period between March 2012 and May 2014, with the goal of testing which type of social safety net transfers worked best to improve household food security and child nutrition outcomes among the ultra-poor in Bangladesh. One trial was conducted in the rural areas of the northwest region of the country ("North"), the other in the south ("South"). In each region, 250 villages were randomly selected across

five subdistricts, and then ten households that met the following criteria were randomly selected from each of the 250 villages: (1) the household was poor (defined as their total consumption expenditure being less than the lower poverty line in Bangladesh), (2) the household had at least one child aged 0-24 months in February-March 2012, when a census was conducted to identify eligible households, and (3) the household was not receiving benefits from other social safety net interventions. Thus, a total of 2500 households in each region – North and South – were part of the study.

Villages from each region were randomly assigned to one of 5 treatment arms. The first three arms were common across both regions: (1) cash only ("Cash"), (2) food only ("Food"), and (3) half cash and half food ("Cash&Food"). The fourth arm was region-specific and consisted of combining either cash or food with complementary programming in the form of nutrition BCC ("Cash+BCC" in the North and "Food+BCC" in the South).[2] Finally, each region had a fifth control arm that did not receive any interventions. The Cash arm beneficiaries received 1500 taka (~USD 19) per household per month for 24 months from May 2012 to April 2014. The Food arm beneficiaries received a food ration that consisted of 30 kilograms (kg) of rice, 2 kg of pulses and 2 litres of micronutrient-fortified cooking oil, the total value of which equaled the cash transfer in the Cash arm. The Cash&Food arm received 750 taka and half the allocation of food. The Cash+BCC and Food+BCC arms received the same transfers as the Cash and Food arms respectively in addition to intensive nutrition BCC interventions. The BCC consisted of three components – (1) weekly group trainings led by a community nutrition worker (CNW), either with beneficiaries only or with beneficiaries and their family members, (2) bimonthly (or more frequent) visits by CNWs to beneficiaries' homes to provide individual counselling, and (3) monthly group meetings between program staff and influential community leaders.

The weekly group trainings were the core component of the BCC. About 10 beneficiaries were part of each group. Trainings took place within 2 km of beneficiaries' homes and lasted approximately one hour. The group trainings solely for beneficiaries occurred on the day of the transfer distribution, once a month. For the remaining group trainings, other household members – particularly mothers-in-law, husbands, and other pregnant or lactating women – were invited to attend along with beneficiaries. The female CNWs leading sessions were engaged by a local NGO specifically for TMRI. CNWs were trained to impart BCC

[2] The number of arms that could be implemented, particularly including BCC, was constrained by budget and logistics of the pilot study, thus we were unable to include a Cash+BCC arm in the South or a Food+BCC arm in the North.

content using a variety of interactive methods including question and answer, flashcards, real-life examples, discussions, practical demonstrations, role-playing, and songs.[3] Sessions were lively, and were viewed by many beneficiaries as welcome opportunities for social interaction given local norms of female seclusion (Roy et al., 2019). Attendance at the group BCC sessions was a soft conditionality in the sense that no beneficiaries were dropped from the study for failing to attend; when a mother missed a session, a CNW would follow up with a home visit to uncover why she missed the session and to convey the missed information. However, attendance was high; beneficiaries attended on average 48 (in the North) to 49 (in the South) group training sessions per year.

Training topics included basic nutrition, control and prevention of micronutrient deficiencies, infant and young child feeding, health care, maternal nutrition, and hygiene. Although the training did not include a specific theme on stimulating children to promote ECD, sessions emphasized the importance of responsive feeding. Messaging included encouraging children to eat by singing songs and telling stories rather than forcing them to eat; feeding infants patiently when introducing complementary foods; letting them eat using their own hands; etc. The nutrition BCC was also framed overall to mothers as promoting behaviors that would help their children's mental growth. Thus, ECD was not the explicit focus of the BCC; however, recommended behaviors included stimulation activities, ECD was described as the motivation for these recommendations, and the BCC overall increased salience on the importance of positive parenting practices.

The twice-per-month home visits were also conducted by CNWs, to observe household-level practice and encourage the adoption of positive nutrition-related behaviors. CNWs conducted additional home visits as needed for individual counseling to beneficiaries or to follow up on missed group sessions. The monthly group meetings with influential community members (such as village heads, religious leaders, school teachers, community elected persons, and local health and family planning staff) were conducted by CNWs and staff from the local NGO, without the beneficiaries present, to explain the purpose and importance of the nutrition BCC and to provide them with the information being conveyed to study participants.

[3] The BCC component was designed specifically for TMRI by the World Food Programme in consultations with the International Food Policy Research Institute and local technical experts. Session materials were derived in part from material developed for Alive & Thrive in Bangladesh, a global nutrition initiative that seeks to save lives, prevent illness, and ensure healthy growth of mothers and children.

3. Data and Estimation Sample

The initial evaluation of the TMRI interventions included three rounds of longitudinal surveys. The baseline survey was conducted in March-April 2012, prior to the start of the intervention in May 2012; a midline survey was conducted in June 2013, and an endline survey was conducted in April 2014. Subsequently, two additional rounds of surveys were conducted to assess the post-program impacts of the TMRI interventions. The first of these was conducted between October 2014 and February 2015, six to ten months after the intervention ended (denoted "6mPP" for "6 months post-program"). The second was conducted between April and May 2018, i.e., four years after the intervention ended (denoted "4yPP" for "4 years post-program").

The baseline, midline, and endline surveys attempted to interview all 5,000 households that were included in TMRI treatment or control groups. At baseline, in each household, the youngest child aged 0-24 months at the time of the census in February-March 2012 was identified as the "index" child. However, as the focus of the original impact evaluation was to assess TMRI's impacts on household food security and child nutrition, ECD and home environment were not measured during baseline, midline, or endline rounds.

The 6mPP and 4yPP rounds explicitly aimed to assess TMRI's post-program impacts on child development and home environment. At 6mPP, two ECD assessments were selected based on the age range of index children (2-4 years old), hypothesized domains of intervention impact, study context, field budget, and field logistics: (1) the Wechsler Preschool and Primary Scale of Intelligence, Third UK edition (WPPSI-III), for a measure of intelligence quotient (IQ), and (2) the Motor Assessment Battery for Children (MABC), for a measure of motor development. For a measure of home environment, the Home Observations for Measurement of the Environment (HOME) inventory was used, along with several questions on time spent stimulating the child in various activities adapted from the Multiple Indicator Cluster Surveys (MICS). To administer these, the index child and the child's mother (or if she was not available, the child's primary female caregiver, referred to as the mother hereafter for simplicity[4]) were requested to go to a village center – usually a school or community club – where tables and chairs were available in a quiet room free of distractions, as many of the assessment items involved simple table-top games using blocks, pictures, and other manipulatives. A trained enumerator first conducted ECD assessments with the child, then administered a short instrument to the child's mother on several topics including the home environment.

[4] At 6mPP, 2.7 percent of caregivers were not biological mothers. At 4yPP, 5 percent of caregivers were not biological mothers.

Due to budget constraints, the 6mPP survey included only a subset of the TMRI intervention arms: in the North, (1) Cash, (2) Cash+BCC, (3) Control; and in the South, (1) Food, (2) Food+BCC, (3) Control. As a result, at 6mPP, we attempted to re-interview only 3,000 out of the initial 5,000 baseline households. Further, because the WPPSI-III was designed for children 30 months or older, the 6mPP sample was restricted to households with index children aged at least 30 months by October 2014.

At 4yPP, a different set of child development assessments was selected based on index children now being approximately 6-8 years old, and accounting for conceptual, contextual, budgetary, and logistical considerations as above. The child assessments included the Raven's Standard Progressive Matrices (RSPM), the Early Grade Reading Assessment (EGRA), the Early Grade Math Assessment (EGMA), and the Strengths and Difficulties Questionnaire (SDQ). For measures of home environment, MICS questions on time spent stimulating the child were re-administered. The 4yPP interviews were conducted as part of a larger household survey, thus all items were administered at or just outside the respondents' homes. The assessments did not require materials aside from booklets with pictures, thus tables and chairs were not needed. Attempts were made by trained enumerators to limit distractions or interruptions during the child assessments, and mothers were separately interviewed on the SDQ and MICS questions.

For our analysis of how TMRI's post-program impacts on child development evolve over time, we consider index children who appeared in both the 6mPP and 4yPP rounds and who have complete data on relevant outcomes. Figure 1 shows the construction of our estimation sample. Of the 3,000 baseline households in the relevant treatment arms, 2,830 were re-interviewed at 6mPP. Of those, 167 were excluded because either the index child's data were missing (for reasons that included the mother or index child not being at home or having migrated, refusal of either to participate, or the child's inability to finish the interview), or the child could not be matched to the mother. This left us with 2,663 index children-mother pairs at 6mPP that had complete information on ECD, HOME inventory and maternal outcomes. Of the 2,663 index children-mother pairs at 6mPP, 170 index children were not reinterviewed at 4yPP (attrition) and were dropped from the sample. Finally, we dropped 113 children due to applying the same sample restrictions as at 6mPP and excluding households with index children whose birthdate at baseline reflected that they had not been within 0-24 months at the time of the census, thus outside the intended age range for index

children. This resulted in a final analysis sample of 2,380 index children that have complete baseline, 6mPP and 4yPP data.[5]

4. Outcomes

4.1 Home environment and maternal outcomes

Home observations for measurement of the environment (HOME)

At 6mPP, we administered the HOME inventory (Caldwell, 1967), which measures the quality and amount of stimulation available to a child in their home environment. The inventory is designed for children 0-3 years and was adapted to Bangladesh. [6] Items included questions from four subscales - learning materials (12 items; the availability of books, toys, writing and drawing material, motor toys etc.), language and academic stimulation (11 items; whether the parents spend time talking to their child, tell the child stories, teach their child the names of colors, basic arithmetic, writing and drawing etc.), variety (4 items; whether the child has a musical instrument, is taken on an outing at least twice a month, needs to eat with their parents at least once a day etc.) and acceptance (2 items; whether the parents hit the child when he/she behaves badly, and what they do when the child hits back). All items were scored 0/1, except one question from the acceptance sub-scale that was scored 0/1/2. Scores on these four subscales were added to create the overall HOME variable (Hamadani et al. 2010), which ranged from 0 to 30. The HOME inventory was not administered as part of the 4yPP survey.

Maternal depression

At 6mPP we administered six questions from the Centers for Epidemiological Studies-Depression (CES-D) scale (Radloff, 1977) which consists of reading statements relating to symptoms of depression and asking respondent women to indicate how many days in the previous week they had experienced the symptom. These six questions are part of the Family Care Indicators assessment (Hamadani et al. 2019) and have previously been found to discriminate between depressed and non-depressed adults (Wolf et al. 2002).

[5] Among these 2,380 children, MABC scores were unavailable for 215 children at 6mPP because of a data entry issue, however other items are complete for these children.

[6] Although the sample of children was 2-4 years old in this round, we administered the items for 0-3 years to all for simplicity in the field.

Statements included “I felt sad”, “I felt lonely”, “I had crying spells”, “I enjoyed life”, “I felt depressed” and “I felt little interest or pleasure in doing things”. “I enjoyed life” was reverse-coded, and responses on all six questions were summed to create a maternal depression scale that ranged from 0 to 42. Higher values on this scale indicate greater levels of depression.

At 4yPP we used the Patient Health Questionnaire with 9 items (PHQ-9) following Kroenke et al. (2001) to screen for depression. Similar to 6mPP, women were asked to indicate how frequently they experienced a set of depressive symptoms in the past week, rating these on a scale of zero (never) to three (nearly every day). Responses on all 9 items were summed which resulted in a PHQ-9 depression score ranging from 0 – 27, depending on the frequency of each occurrence. Again, higher values indicated greater levels of depression, with depression severity ranging from minimal depression (1-4) to mild (5-9), moderate (10-14), moderately severe (15-19), and severe depression (20-27).

Time spent stimulating the child

At both 6mPP and 4yPP we asked mothers about stimulation activities they had performed over the past week with the index child, based on questions adapted from the MICS surveys. Specifically, we recorded the amount of time they or other household members aged 16 and over had spent telling stories to the child, singing songs to or with the child, reading books or looking at pictures with the child, taking the child outside the homestead, playing with the child, and naming, counting or drawing things with the child. The reported amounts of time in minutes were summed across activities to create two indicators – one for the total amount of time the mother spent stimulating the child, the second for the total amount of time any adult in the household aged 16 or older spent stimulating the child.

4.2 Child development outcomes

4.2.1 Outcomes at 6 months post-program

At 6mPP, we administered two separate tests of early childhood development to the children in our sample. The first was the Wechsler Preschool and Primary Scale of Intelligence – Third UK edition, or WPPSI-III, used to assess children’s IQ. The second was an assessment of motor skills consisting of selected tests from the Motor Assessment Battery for Children or MABC. These tests differ by the age of the child,

and all had been previously adapted for and shown to have validity in assessing Bangladeshi children (Hamadani et al. 2013; Hamadani et al. 2014).[7]

Children aged 2-4 years at the time of the 6mPP survey (henceforth "younger children") were administered the WPPSI-III tests on two domains, verbal and performance. For the verbal domain, 2 tests were administered on receptive vocabulary and information, and for the performance domain, 2 tests were administered on block design and object assembly. Raw scores on each test are converted to scaled scores based on the age of the child. The sum of scaled scores within each domain are then converted to IQ equivalents using age-dependent conversion tables, yielding a verbal IQ (VIQ), performance IQ (PIQ) and full-scale IQ (FSIQ) score for each child. The full-scale IQ is the sum of verbal and performance IQ scores.

Children aged four years and above (henceforth "older children") were administered the WPPSI-III tests across three domains, verbal, performance, and performance speed. A total of seven tests were administered: information, vocabulary and comprehension (verbal domain), block design, matrix reasoning and picture completion (performance domain) and coding (performance speed domain). Once again, raw scores on these tests are converted to age-specific scaled scores, and the sum of scaled scores across the set of tests within a domain is then converted to an IQ score on that domain. Since only one test was administered under the performance speed domain, we do not convert that score into an IQ equivalent, and instead report only the VIQ, PIQ and FSIQ scores.

Children above the age of three years at the 6mPP survey were administered three tests on the MABC: the number of seconds taken to thread a certain number of beads ("threading beads"), the number of throws out of 10 that the child was able to catch a beanbag tossed to them ("catching a bean bag"), and the number of steps the child could walk with their heels raised ("walking heels raised"). Other subscales were not administered, as previous work had found them to be too easy or too difficult for Bangladeshi children in this age group (Hamadani et al. 2013). Raw scores on each test are converted to age-specific standard scores, and then summed across tests to arrive at the overall score on the MABC. The highest possible raw score on the walking heels raised test was 15. Since almost all children in our sample scored 15 on this test, we drop this due to apparent ceiling effects, and report the scores on the 2-item MABC.

[7] At 6mPP, most index children's ages ranged from 2.5 years to 4 years, with a small number of children who were just over 5 years, given that eligibility was based on being 0-24 months as of February-March 2012, and the 6mPP survey round spanned from October 2014 to February 2015.

4.2.2 Outcomes at 4 years post-program

At 4yPP we administered four separate tests of child development: The Raven's Standard Progressive Matrices (RSPM), the Early Grade Reading Assessment (EGRA) for Grade 1, the Early Grade Math Assessment (EGMA) for Grade 1, and the Strengths and Difficulties Questionnaire (SDQ). These assessments focused on children's age-relevant abstract reasoning, foundational literacy and math skills, and behavioral strengths and difficulties. As part of the 4yPP survey preparations, the EGRA was modified for age-appropriate difficulty within the study context by a Bangladeshi education expert, and all test materials and instructions were translated to Bangla.[8] All children in the 4yPP sample were administered the same tests, without additional distinction within the 6-8-year-old age range.

The Raven's Standard Progressive Matrices is a pattern recognition test that attempts to measure abstract reasoning and is considered to be an estimate of fluid intelligence. It consists of several sets of multiple-choice questions, which display different geometrical patterns with missing pieces and ask the respondent to select among various pieces the one that matches the best. While there are different sets of Matrices for different age ranges and ability levels, we administered the first two sets A and B of the Standard Progressive Matrices with 12 items each. Research has shown that abbreviated forms of the RSPM produce equally accurate results than the full 60-item test (Bilker et al., 2012). The total raw score was then simply calculated as the sum of correct responses across all 24 items, yielding a score ranging from 0-24.

The EGRA (Dubeck and Gove, 2015) measures reading acquisition skills in early primary graders and is a comprehensive toolkit with several subtasks that are separately administered. We administered four core subtasks: listening comprehension – where children are scored based on the number of correct answers to three comprehension questions; letter recognition – where the children's score is calculated as the number of correct letter sounds read per minute; unfamiliar "non-word" reading – where the children's score is calculated as the number of correct non-words read per minute; and oral reading and comprehension – where two scores are calculated: the number of words read correctly per minute and how many of the 4 reading comprehension questions were answered correctly. Given that these subtasks

[8] For example, in the EGRA, a sentence in English considered of "Grade 1" difficulty was often more or less difficult when directly translated to Bangla, thus the content of many of the listening and reading comprehension passages needed to be modified. Similarly, some combinations of letter sounds in English had no corresponding letters in Bangla, thus were modified to be of comparable difficulty rather than direct translations.

increase in difficulty as they progress, missing values in later subtasks were recoded to zero, under the assumption that children did not start them due to their difficulty and would have scored zero on them if they had attempted them.

The EGMA measures foundational mathematics knowledge and is administered through separate subtasks, analogous to EGRA. The four subtasks that we administered are: number identification – the score being the number of correctly identified numbers shown per minute; counting – where the score is composed of how many symbols in the shown pictures are counted per minute, and how many corresponding numbers children can point out correctly; number discrimination – where children are shown several rows of numbers and the score is calculated based on how many of the smallest and largest numbers per row children can identify per minute; and addition and subtraction – where children are scored based on the number of correct answers to 5 basic addition and 5 basic subtraction questions. The addition and subtraction subtask is the only untimed EGMA subtask we administered.

Lastly, all mothers were administered the Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997), to be answered about the index child. The SDQ is a commonly used behavioral and mental health screening questionnaire for children and adolescents. Mothers were read 25 descriptive statements about their child's behavior over the previous 6 months and asked if these statements were "not true", "somewhat true", or "certainly true". The 25 items were grouped into 5 separate scales: emotional problems, conduct problems, hyperactivity, peer problems, and prosocial. Statements included "considerate of other people's feelings", "restless, overactive, cannot stay still for long", or "often lies or cheats". Positive statements were then reverse-coded, and all scales except the prosocial scale were summed up to a total SDQ raw score.[9] This Total Difficulties Score ranges from 0-40, with higher scores indicating greater behavioral difficulties.

4.3 Age standardization of child development outcomes

Since raw test scores naturally increase with age, we standardize them in order to remove the age effect and allow for within-sample comparisons. That is, we aim at comparing outcomes across all children in our sample. We follow Attanasio et al. (2020) and use non-parametric regression methods to compute both the empirical age-conditional mean and age-conditional standard deviation for each item and age in

[9] The prosocial scale is semantically different and commonly excluded from the total score. For more information, see https://sdqscore.org/.

our sample. These age-conditional means and standard deviations are then used to calculate age-adjusted z-scores for each outcome. As a result, these z-scores then have sample means of zero and sample standard deviations of one, with the age effect being removed.[10]

For both the EGRA and EGMA, the literature supports analyzing the different subtasks separately (Graham and Kelly, 2018; Gove and Wetterberg, 2011; Ketterlin-Geller et al., 2018) – as they are intended to capture distinct domains – rather than combining the different subtasks into a single composite index.[11] Our main analysis on EGRA and EGMA thus analyzes impacts on each subtask indicator distinctly. However, as a robustness check, we also assess impacts on a composite index for each. Specifically, we take the age-standardized subtasks of the EGRA and EGMA and aggregate them into combined EGRA and EGMA summary indices following Kling, Liebman, and Katz (2007) (Appendix A.5). Doing so complicates the interpretability of the outcomes but reduces the number of outcomes assessed.

5. Empirical Specification

For each of the home environment, mother's wellbeing, or child development outcomes Y described above, we estimate the following regression for individual i in household h in village v using standard ordinary least squares (OLS):

$$Y_{ihv} = \alpha + \beta_1 Transfer_v + \beta_2 TransferBCC_v + X'_{ihv}\gamma + \varepsilon_{ihv} \qquad (1)$$

We do so separately for each region, North and South. $Transfer_v$ is an indicator that equals 1 if the village was assigned to the Cash (North) or Food (South) treatment and β_1 is the intent-to-treat (ITT) estimate of being assigned to a transfer arm relative to the control group. $TransferBCC_v$ is an indicator that equals 1 if the village was assigned to the Cash+BCC (North) or Food+BCC (South) treatment and β_2 is the ITT

[10] Cf. Appendix B in Attanasio et al. (2020) for a more detailed, formal description of this standardization process.

[11] Graham and Kelly (2018) review 18 EGRA impact studies and present all results by subtask. Gove and Wetterberg's (2011) comprehensive book on how the EGRA has been used in several countries around the world discusses the subtasks but not any composite. Ketterlin-Geller et al. (2018) state regarding the EGMA, "The developers of the EGMA recommend that results from each subtest be reported individually as subscores (RTI International, 2014), as opposed to aggregating scores from multiple subtests to form a composite or total score. This is the most common practice for reporting EGMA results (c.f., Brombacher et al., 2015; Piper & Mugenda, 2014; Torrente et al., 2011)."

estimate of being assigned to a transfer plus BCC arm relative to the control group. X_{ihv} is a vector of individual or household-level controls, included to improve the precision of our estimates, and γ is a vector of associated coefficients. Standard errors in all regressions are adjusted for clustering at the village level, which is the level of randomization. To test whether the coefficients β_1 and β_2 are statistically significantly different from one another, we perform Wald tests of equality. We estimate equation 1 for the combined sample of boys and girls and disaggregated by sex.

For regressions using woman-level reported outcomes (such as the home environment, maternal depression and time spent stimulating the child), we control for the baseline values of the respondent woman's age and education, household size and per capita monthly household expenditure. For regressions using child-level outcomes on the IQ or motor tests, we control for baseline values of mother's age and education, a dummy for the father being in the household at baseline, a dummy for the child being stunted (i.e., having a height-for-age z-score of less than 2), household size and per capita monthly household expenditure. All regressions additionally control for child sex and child age at the time of the survey, except regressions that disaggregate by sex, where the control for child sex is removed. Missing values in control variables, if any, are replaced with 0s, and dummies for missing values are added.[12]

In addition to using OLS, at 4yPP we also use a series of alternative models such as Tobit, Probit, and ordered Probit as robustness checks for impact estimations on child development outcomes. These models use an analogous specification to (1) and are selected to better account for non-normal distributions. The specific selection of the models depends on the respective distribution of the outcome. For instance, children may find some tests at 4yPP more difficult than others, leading to floor effects due to the large number of zero scores for these tests. In this example, a Tobit regression would provide more efficient coefficient estimates than OLS.

[12] Three dummy variables are added for missing values of control variables, corresponding to 15 missing values for mother's age at baseline, 12 missing values for mother's education at baseline, and 7 missing values for whether the child was stunted at baseline.

6. Results

6.1 Attrition

Attrition from the 3000 index children and mothers at baseline to the final estimation sample of 2380 at 4yPP was 21 percent, or about 3.4 percent per year between 2012 and 2018. Attrition does not systematically differ across treatment arms in either the North or the South (Appendix Table A.1a). Tests of joint orthogonality for baseline child, mother, and household characteristics do not show any significant differences between the Cash and Control, Cash+BCC and Control, or Cash and Cash+BCC arms in the North (Appendix Table A.1b); the test of joint orthogonality between the Food and the Control arm in the South has a significant p-value of 0.04 (Appendix Table A.1c), suggesting a possibility of nonrandom selection into the estimation sample in the South.

Appendix Tables A.1d-A.1e and A.1f-A.1g present the same orthogonality tests by region for the sample of girl and boy children respectively. In the North, the F-test of differences in baseline characteristics for girls in the Cash+BCC versus the Control arms has a marginally significant p-value of 0.08 (Table A1a). In the South, the test of differences between baseline characteristics of the sample in the Food and Control arms is also significant with a p-value of 0.04 (Table A1d).

Overall, the validity of the TMRI experimental design appears largely intact in the final 6mPP and 4yPP estimation sample. Our inclusion of baseline covariates helps adjust for remaining differences across arms.

6.2 Main results at 6 months post-program

6.2.1 Home environment outcomes

Tables 1a and 1b present the main impact estimates of TMRI on home environment outcomes at 6mPP, for the North and South. In the North (Table 1a), the Cash treatment arm has a positive impact on the HOME inventory score of 1.3 points, or 11 percent increase relative to the control mean; however, it does not have a significant impact on maternal depressive symptoms or time spent stimulating the child. The Cash+BCC treatment, on the other hand, causes a 2.8 point increase in the HOME score (23 percent increase relative to the control mean) and a 1.92 point decline in maternal depressive symptoms (20 percent decline relative to the control mean). Cash+BCC in the North also increases the amount of time the mother or any adult in the household spent stimulating the child in the previous week by 100 minutes and 128 minutes respectively, which represent a 32 and 34 percent increase over the 6mPP control means,

respectively. The difference in coefficients between Cash and Cash+BCC is significant for the HOME inventory score and maternal depressive symptoms, though not for time spent stimulating the child.

In the South, impacts are more muted (Table 1b). The Food treatment does not have a significant impact on any of the four outcomes. The Food+BCC treatment causes an increase of 1.5 points in the HOME inventory score (11 percent increase relative to the control mean) but does not significantly affect maternal depressive symptoms or time spent stimulating the child. The difference in coefficients between Food and Food+BCC is significant only for the HOME inventory score.

Tables 1c and 1d disaggregate the 6mPP impacts on home environment outcomes by child sex. In the North (Table 1c), the Cash arm significantly increases the HOME environment score for both girls and boys. The Cash+BCC treatment significantly increases the HOME inventory score for both girls and boys by similar amounts, significantly reduces maternal depressive symptoms only for mothers of boys, significantly increases time spent by the mother stimulating girls in the previous week, and significantly increases the time spent by all adults on stimulating girls. The Cash+BCC treatment has no impact on time spent by the mother or other adults in stimulating boys. The difference between the Cash and Cash+BCC treatments is significant for the HOME inventory score for both boys and girls.

In the South (Table 1d), impacts are smaller and less significant. When disaggregated by child sex, the Food treatment does not have a significant impact on any of the four outcomes. The Food+BCC treatment positively impacts the HOME inventory score of both boys and girls. It also increases the time spent by any adult in the household on stimulating girls by 82.78 minutes in the previous week, which amounts to a 20 percent increase over the control mean. The difference between the Food and Food+BCC treatments is significant for the HOME inventory score for both boys and girls.

6.2.2 Child development outcomes

Tables 2a and 2b present the impacts of TMRI on age-standardized child development outcomes for younger and older children, by region. Overall, we see little evidence of impact of the Transfer or Transfer+BCC treatments in either region. In the North (Table 2a), the Cash treatment has no impact on IQ outcomes for either younger or older children. The Cash+BCC treatment causes a decline in the standardized score on the performance IQ of younger children of 0.17, and an increase in the standardized scores on verbal and full-scale IQs for older children of 0.23 and 0.17 respectively. We do not see any impact of either treatment on motor scores of younger or older children. In the South (Table 2b), the only significant impact is that of the Food+BCC treatment on standardized motor scores of older children,

causing a decrease of 0.24 over a control mean of 0.15.

To assess robustness, we conduct analogous estimation on the raw non-standardized IQ and motor scores (Appendix Tables A.2a and A.2b), as well as on the item-wise tests under each IQ domain and on the MAB (Tables A.3a-d). The results are qualitatively very similar to those described above. In the north, we see some positive impacts of the Cash+BCC treatment on scores on the information and comprehension items of the Verbal IQ domain (Table A.3c). Other than that, there are very few impacts of the treatments on either composite or item-wise scores, particularly in the South.

We next disaggregate 6mPP ECD impacts by child sex. For younger children in the North (Table 2c), the Cash+BCC arm causes a decrease in the performance IQ and 2-item MAB test scores only for girls. For older children in the North, the Cash+BCC increases verbal IQ among boys; all other impacts are insignificant. For younger children in the South (Table 2d), the Food arm causes a weakly significant increase in the standardized verbal IQ for girls of 0.30, while the Food+BCC treatment causes a weakly significant decline in the standardized verbal IQ for boys of 0.27. For older children in the South, the Food+BCC causes a weakly significant decrease in the 2-item MAB for girls. Overall, most impacts are statistically insignificant or only marginally significant at the 10 percent level, and the few that are significant do not point to any clear pattern of impact by child sex.

6.2.3 Summary of 6-month post-program results

Before turning to the 4yPP results, we summarize the 6mPP results. In the North, both Cash and Cash+BCC cause significant increases in the HOME inventory, although the impacts of Cash+BCC are significantly larger; only the Cash+BCC causes significant reductions in maternal depressive symptoms and significant increases in time spent stimulating the child. When disaggregated by sex, improvements from Cash and Cash+BCC in the HOME inventory are significant for both boys and girls, impacts of Cash+BCC on maternal depressive symptoms are driven by mothers of boys, and impacts of Cash+BCC on time spent stimulating children are driven by girls. In the South, the Food+BCC treatment increases the HOME inventory score, but no other impacts are significant. When separated by child sex, improvements from Food+BCC in the HOME inventory appear for both boys and girls, while small increases in time spent stimulating the child emerge for girls. Overall, the impacts on home environment are stronger in the North than the South, and in all cases except that of maternal depressive symptoms in the North, impacts tend to appear stronger for girls than boys. Given the prevalence of son preference in the region, it is worth noting that age-relevant mean values in the Control arm for these outcomes do not seem to differ substantially by child sex, though it does appear that mothers and other adults spend more time on stimulation activities with

boys. Thus, the impacts we find do not reflect an obvious widening or narrowing of an existing gender gap among children ages 2 to 4 years old.

Despite the positive and large impacts on home environment outcomes in the North, we do not see strong impacts on child IQ or motor skills for the younger children and find a significant increase only in the verbal and full-scale IQ scores of older children in the North. Disaggregating the ECD results by child sex suggests that the impacts on IQ among older children in the North were driven by boys, even though larger improvements tended to be seen in the home environment outcomes for girls. In the South, perhaps less surprisingly given the limited impact on home environment scores, there are few impacts on younger or older children's IQ or motor scores; sex disaggregations suggest weak and inconsistent patterns by sex.

There are several reasons why we might not see a strong pattern of impacts on ECD outcomes at 6mPP. First, it is possible that the tests administered were not well suited to the children in our sample, that is, they were too hard, too easy, not well understood, or not well-suited to capture intended domains in the study context. Since the tests were adapted and validated by in-country child psychologists for similar-age children in the Bangladeshi context, and since the interviewers were trained extensively on how to conduct these tests in an appropriate and standardized way, this explanation seems unlikely to be the main driver of the lack of impacts.[13] We also conducted simple diagnostics around the validity of the IQ and motor scores against select baseline characteristics and found that the tests perform largely as expected[14]. A second possibility is that there were simply no impacts of TMRI on child development. Given the prior findings on child diets and nutrition during the intervention, the strong 6mPP impacts on home environment and stimulation in the North, and the links in the literature between these measures and child development indicators including in Bangladesh, this would be surprising, although not impossible.

[13] We note that it is possible that the samples of Bangladeshi children for which the tests were validated differ from the very poor rural children in our study sample.

[14] These tests consisted of checks of monotonicity of the scores on IQ tests by mother and father education, household consumption expenditure and child stunting at baseline. A priori, we would expect positive associations with parental education and per capita household consumption expenditure, and negative with child stunting, which is what we found (Appendix Figure A.1). Motor test scores are more closely tied to child nutrition status, so we tested their association with per capita household consumption expenditure and child stunting, wasting and underweight (Appendix Figure A.2). Except for household expenditure, all other associations are in the direction we would expect.

An alternative possibility is that the changes in the home environment and child stimulation take longer to manifest themselves in improvements in test scores. It is with this lens that we turn to the 4yPP results.

6.3 Main results at 4 years post-program

6.3.1 Home environment outcomes

To assess how TMRI's home environment impacts at 6mPP evolve over time we estimate impacts on the same sample 4 years post-program. Tables 3a and 3b show impacts on maternal depressive symptoms and time spent stimulating the child in the North and the South, respectively. In the North, we find significant reductions in maternal depressive symptoms at 4yPP in both Cash and Cash+BCC arms relative to the control group: 14 percent and 22 percent, respectively. However, only Cash+BCC significantly increases the amount of time the mother or any adult in the household spent stimulating the child in the previous week. On average, children in Cash+BCC households were stimulated 63 minutes and 55 minutes longer than children in the control group, respectively. In the South, neither Food nor Food+BCC have any significant impact on any home environment outcomes at 4yPP.

We next disaggregate these results by sex. The results in the North (Table 3c) show that the impacts on maternal depressive symptoms and time spent stimulating the child are exclusively driven by boys. None of the impacts in the sample of girls are significantly different from zero. In the South (Table 3d), the impacts of Food or Food+BCC on all home environment outcomes are statistically insignificant.

6.3.2 Child development outcomes

We next assess TMRI's impacts on 4yPP child development outcomes for the same sample as at 6mPP. Table 4a shows that in the North, Cash significantly improves the Raven score (0.23 SD) as well as two EGMA subtasks – number identification (0.16 SD) and counting (0.14 SD) – but has no significant impacts on other outcomes. The only significant impact from Cash+BCC in the North is a 0.23 SD improvement in the SDQ score (lower SDQ scores indicate less behavioral difficulties). No evidence of impact is seen on any of the EGRA reading outcomes from Cash or Cash+BCC in the North.

In the South (Table 4b), Food+BCC treatment causes a 0.15 SD weakly significant decline ($p<0.10$) in the EGRA non-word reading score, and Food leads to a 0.15 SD weakly significant decline in the EGRA reading comprehension score. All other impacts are statistically insignificant.

We conduct several robustness checks. Estimating the same model specification using the raw, non-standardized CD scores shows similar results (Appendix Tables A.4a and A.4b), except that the negative point estimate on EGRA reading comprehension in the South is no longer significant. Aggregating EGRA and EGMA subtests to summary indices (Appendix Tables A.5a and A5b) shows no significant impacts on the summary indices in either region. Using alternative estimation models for each outcome that are more sensitive to the distribution of the respective raw scores – such as Tobit or Probit – do not lead to substantially different results than OLS (Appendix Tables A.6a and A.6b).

We next disaggregate the 4yPP CD results by child sex. In the North (Table 4c), we see that the significant impacts found in the pooled sample in the North (Table 4a) are largely driven by boys. In particular, the significant improvements in EGMA number identification and counting from Cash as well as the significant improvements in SDQ from Cash+BCC are concentrated among boys; however, Cash leads to weakly significant positive impacts on the Raven score among both boys and girls. In the South (Table 4d), the pattern by sex is less clear, although program impacts among both boys and girls are either declines or insignificant. In particular, Food+BCC shows a weakly significant negative impact on EGRA letter recognition among girls; among boys, Food shows negative impacts on EGRA reading comprehension and EGMA number discrimination, while Food+BCC shows a weakly significant worsening of the SDQ score.

As a robustness check, we assess impacts by child sex on the composite indices for EGRA and EGMA (Appendix Table A.5c and Appendix Table A.5d). We find that Cash in the North shows a significant positive impact on the EGMA composite index among boys, while Food in the South shows a weakly significant negative impact on the EGRA composite index among boys. There are no significant impacts on the indices among girls in either region.

6.3.3 Summary of 4-year post-program results

Overall, at 4yPP we find positive impacts on home environment outcomes in the North; both Cash and Cash+BCC lead to significant reductions in maternal depressive symptoms, and Cash+BCC additionally increases time spent stimulating the child. Results indicate that these improvements in home environment in the North are concentrated among boys, with no significant impacts among girls. There are also no significant impacts on home environment from Food or Food+BCC in the South among boys or girls. Notably, in both regions, girls' home environment outcomes tend to appear better than boys' in the control group, suggesting that program impacts are narrowing the gender gap rather than widening it.

In terms of child development outcomes, we see some positive impacts in the North: Cash improves the Raven score and EGMA number identification, while Cash+BCC improves the SDQ score. Again, child development impacts in the North appear to be driven by boys, although the improvement in Raven scores appears among both boys and girls. By contrast, in the South, there is little evidence of improvements in child development – and suggestive evidence of declines, particularly in EGRA subtasks. Patterns in child development impacts by sex are less clear in the South, though declines tend to appear larger for boys. Again, in both regions, girls' child development outcomes generally appear higher than boys' in the control group, indicating that program impacts (where significant) are also narrowing this gender gap rather than widening it.

7. Discussion and conclusion

We conduct a post-program assessment of a transfer program's impacts on children's home environment and child development – exploring the role of complementary programming and context, as well as how impacts differ for boys and girls, and how impacts evolve over time. We find that in the Northern region of rural Bangladesh, Cash or Cash+BCC leads to significant post-program improvements in measures of home environment at both 6mPP and 4yPP – including the HOME inventory, maternal depressive symptoms, and time spent stimulating the child. However, these translate to limited improvements in post-program child development. At 6mPP, we see a small positive impact from Cash+BCC on verbal IQ among older children in our target range (about 4 years old); at 4yPP, among 6- to 8-year-olds, we see that Cash improves the Raven score and EGMA number identification, and Cash+BCC improves the SDQ score, but there are few other impacts including on the EGRA. While impacts on home environment in the North appear larger among girls at 6mPP, home environment improvements are driven by boys at 4yPP. The few improvements in child development from Cash or Cash+BCC at 4yPP also appear driven by boys. In the Southern region by contrast, we see few impacts on home environment or child development, either at 6mPP or 4yPP. In fact, where impacts are (weakly) significant, they indicate declines in child development from Food or Food+BCC, and patterns by sex are unclear.

There are several interesting observations from our analysis. First, our results are promising for the potential of cash transfers to improve children's home environment beyond the end of programming – particularly with the addition of complementary programming for mothers. Impacts from Cash+BCC in the North are sustained at 4yPP across children's mothers' psychological well-being and time spent

stimulating the child. Although the home environment impacts of Cash+BCC do not differ significantly from impacts of Cash at 4yPP, they tend to be larger, suggesting a potentially important role for BCC – particularly in terms of time spent stimulating the child, which does not rely directly on a household's resources. Second, 4yPP impact estimates indicate that the addition of BCC – and resulting increase in parents' time spent stimulating children, along with improved diets in the first 1000 days – may particularly benefit children's behavioral outcomes by the time they are 6 to 8 years old. Although impacts of Cash+BCC on behavioral indicators do not significantly differ from that of Cash, they are larger. Behavioral outcomes are of interest, as they may relate to the non-cognitive domains that the literature increasingly shows are important for long-term outcomes (Heckman et al. 2006). Third, Cash by itself shows impacts on measures of cognitive development in the North by 4yPP, particularly for boys. This finding suggests that not all improvements in child development require complementary programming. Fourth, findings suggest that cash transfers may better promote post-program improvements in children's home environment than food transfers, although this inference is complicated by our study design's indeterminacy on transfer modality vs. context. In particular, we cannot disentangle the role of cash vs. food from North vs. South. Fifth, results indicate that an intervention's impacts on child development may evolve over time. Although our comparisons across rounds are complicated by the necessity of using different instruments for different age ranges, with different relevant developmental domains, findings suggest that Cash and Cash+BCC in the North show slightly stronger improvements in child development at 4yPP than at 6mPP. This pattern over time is consistent with other evidence on early childhood interventions, reflecting that some short-term benefits can fade out even as other benefits emerge over time (Heckman, 2008). Finally, disaggregations suggest that the impacts of transfer programs on home environment and child development can differ by sex. In our study setting, the effects of Cash and Cash+BCC in the North on home environment and child development appear driven by boys at 4yPP.

Several elements of our findings raise questions as well. First, the results we present cannot conclusively explain the link (or lack thereof) between impacts on home environment and child development. Most likely, these relationships are interlinked with other pathways that we do not (and in many cases cannot, given the data we have) analyze here. For example, we do not have detailed information on children's schooling, but by ages 6 to 8, this may be an important factor in child development outcomes; we do not have detailed information on books, toys, and other stimulation materials available to the child at 4yPP; and we do not analyze whether there are sustained impacts on child diets or health investments beyond the end of programming. Future research to better understand potential pathways between transfers and child development across ages would be useful. Second, we find patterns by sex but cannot explain what

underlies them. Given that our control means indicate that girls have slightly higher outcomes on home environment and child development at 4yPP in the North, the concentration of positive impacts among boys could reflect parents narrowing a gender gap, however it is less clear why even reductions in maternal depressive symptoms would be concentrated among mothers of boys, and the pattern is interesting in light of prevailing son preference in the region. Third, our results indicate that the treatment arms in the South may contribute to possible declines in child development outcomes, however it is not clear how or why these might occur. In a companion paper (Ahmed et al., 2020), we find that 4yPP poverty impacts from a Cash arm in the North are stronger than those from the same Cash arm in the South, and that the control group appears to "catch up" more during the post-program period in the South than in the North, suggesting that there are important contextual differences between the regions. Better understanding is thus needed of how contextual factors shape the relationship between transfers and child development. We view these issues as important priorities for future work assessing the impacts of transfer programs on human capital, ultimately informing how to design transfer programs to effectively break the intergenerational transmission of poverty.

References

Ahmed, Akhter, Hidrobo, Melissa, Hoddinott, John F., Koch, Bastien, Roy, Shalini; and Tauseef, Salauddin. 2020. Social protection and sustainable poverty reduction: Experimental evidence from Bangladesh. IFPRI Discussion Paper 1988. Washington, DC: International Food Policy Research Institute (IFPRI).

Ahmed, Akhter; Hoddinott, John F.; and Roy, Shalini. 2019. Food transfers, cash transfers, behavior change communication and child nutrition: Evidence from Bangladesh. IFPRI Discussion Paper 1868. Washington, DC: International Food Policy Research Institute (IFPRI).

Ahmed, Akhter U. & del Ninno, Carlo, 2002. "The Food for Education program in Bangladesh," FCND discussion paper 138, International Food Policy Research Institute (IFPRI).

Arriagada, Ana-Maria, Jonathan Perry, Laura Rawlings, Julieta Trias, and Melissa Zumaeta. 2018. "Promoting Early Childhood Development through Combining Cash Transfers and Parenting Programs." 8670. *World Bank Policy Research Working Paper*. Washington, D.C. https://doi.org/10.1596/1813-9450-8670.

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. https://doi.org/10.1257/aer.20150183

Attanasio, Orazio, Sarah Cattan, and Costas Meghir. 2021. "Early Childhood Development, Human Capital and Poverty." 29362. *National Bureau of Economic Research*. Cambridge, MA. https://doi.org/10.3386/w29362.

Baird, Sarah, Francisco Ferreira, Berk Özler, and Michael Woolcock. 2014. "Conditional, Unconditional and Everything in between: A Systematic Review of the Effects of Cash Transfer Programmes on Schooling Outcomes." *Journal of Development Effectiveness* 6 (1).

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, Tanja Schmidt, and Luca Pellerano. 2016. "Cash Transfers: What Does the Evidence Say? A Rigorous Review of Programme Impact and of the Role of Design and Implementation Features." London.

Beegle, Kathleen, Aline Coudouel, and Emma Monsalve. 2018. "Realizing the Full Potential of Social Safety Nets in Africa." *Africa Development Forum Series*. Africa Development Forum Series. Washington, D.C.

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of Abbreviated Nine-Item Forms of the Raven's Standard Progressive Matrices Test. Assessment, 19(3), 354–369. https://doi.org/10.1177/1073191112446655

Black, M. M., Walker, S. P., Fernald, L., Andersen, C. T., DiGirolamo, A. M., Lu, C., McCoy, D. C., Fink, G., Shawar, Y. R., Shiffman, J., Devercelli, A. E., Wodon, Q. T., Vargas-Barón, E., Grantham-McGregor, S., & Lancet Early Childhood Development Series Steering Committee (2017). Early childhood development coming of age: science through the life course. Lancet (London, England), 389(10064), 77–90. https://doi.org/10.1016/S0140-6736(16)31389-7

Brombacher, A., Bulat, J., King, S., Kochetkova, K., and Nordstrum, L. (2015). National Assessment Survey of Learning Achievement at Grade 2: Results for early grade reading and mathematics in Zambia. Research Triangle Park, NC: RTI International.

Caldwell B.M. (1967) Descriptive evaluations of child development and of developmental settings. Pediatrics 40, 46–54.

Carneiro, Pedro, Lucy Kraftman, Giacomo Mason, Lucie Moore, Imran Rasul, and Molly Scott. 2021. "The Impacts of a Multifaceted Prenatal Intervention on Human Capital Accumulation in Early Life." American Economic Review, 111 (8): 2506-49.

Dubeck, Margaret M. and Amber Gove. 2015. The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations, International Journal of Educational Development, Volume 40, Pages 315-322, ISSN 0738-0593, https://doi.org/10.1016/j.ijedudev.2014.11.004.

Engle, Patrice L et al. 2011. Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. The Lancet, Volume 378, Issue 9799, 1339 - 1353.

Fiszbein, A, N Schady, F H G Ferreira, M Grosh, N Kelleher, P Olinto, and E Skoufias. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington D.C.: The World Bank.

Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: a research note." *Journal of child psychology and psychiatry*, *38*(5), pp.581-586.

Gove, A., & Wetterberg, A. (Eds.) (2011). The Early Grade Reading Assessment: Applications and interventions to improve basic literacy. RTI Press. RTI Press Book No. BK-0007-1109 https://doi.org/10.3768/rtipress.2011.bk.0007.1109

Graham, Jimmy; Kelly, Sean. 2018. How Effective Are Early Grade Reading Interventions? : A Review of the Evidence. Policy Research Working Paper;No. 8292. World Bank, Washington, DC. © World Bank.

Grantham-McGregor, Sally et al. 2007. Developmental potential in the first 5 years for children in developing countries. The Lancet, Volume 369, Issue 9555, 60 - 70.

Hahn, Youjin, Asadul Islam, Kanti Nuzhat, Russell Smyth, and Hee-Seung Yang. 2018. Education, Marriage, and Fertility: Long-Term Evidence from a Female Stipend Program in Bangladesh. Economic Development and Cultural Change Volume 66, Number 2.

Hamadani JD, Tofail F, Hilaly A, Huda SN, Engle P, Grantham-McGregor SM. 2010. Use of family care indicators and their relationship with child development in Bangladesh. J Health Popul Nutr. 28(1):23-33. doi:10.3329/jhpn.v28i1.4520

Hamadani JD, Tofail F, Cole T, Grantham-McGregor S. 2013. The relation between age of attainment of motor milestones and future cognitive and motor development in Bangladeshi children. Matern Child Nutr. Jan;9 Suppl 1(Suppl 1):89-104. doi: 10.1111/mcn.12020. PMID: 23167587; PMCID: PMC6860901.

Hamadani JD, Tofail F, Huda SN, Alam DS, Ridout DA, Attanasio O, Grantham-McGregor SM. 2014. Cognitive deficit and poverty in the first 5 years of childhood in Bangladesh. Pediatrics. Oct;134(4):e1001-8. doi: 10.1542/peds.2014-0694. PMID: 25266433.

Hamadani, Jena D et al. 2019. Integrating an early childhood development programme into Bangladeshi primary health-care services: an open-label, cluster-randomised controlled trial. The Lancet Global Health, Volume 7, Issue 3, e366 - e375.

Heckman, James J., Jora Stixrud, Sergio Urzua. 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. Journal of Labor Economics Volume 24, Number 3

Heckman, James J. 2008. Schools, Skills, and Synapses. NBER Working Paper 14064.

Hidrobo, Melissa, John Hoddinott, Neha Kumar, and Meghan Olivier. 2018. "Social Protection, Food Security, and Asset Formation." *World Development* 101. https://doi.org/10.1016/j.worlddev.2017.08.014.

Ivashchenko et al. 2018. The state of social safety nets 2018 (English). Washington, D.C. : World Bank Group

Ketterlin-Geller, Leanne R., Perry, Lindsey, Platas, Linda M. & Sitbakhan, Yasmin (2018). Aligning test scoring procedures with test uses of the early grade mathematics assessment: A balancing act. Global Education Review, 5 (3), 143-164.

Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83–119.

Kroenke, K., Spitzer, R.L. and Williams, J.B.W. (2001), The PHQ-9. Journal of General Internal Medicine, 16: 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Manley J, Balarajan Y, Malm S, et al. 2020. Cash transfers and child nutritional outcomes: a systematic review and metaanalysis. BMJ Global Health. 5:e003621. doi:10.1136/bmjgh-2020-003621

Piper, B., & Mugenda, A. (2014). The Primary Math and Reading (PRIMR) Initiative: Endline Impact Evaluation. Research Triangle Park, NC: RTI International.

Premand, Patrick, and Oumar Barry. 2020. "Behavioral Change Promotion, Cash Transfers and Early Childhood Development : Experimental Evidence from a Government Program in a Low-Income Setting." 9368. *World Bank Policy Research Working Paper*. Washington, D.C. https://doi.org/10.1596/1813-9450-9368.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. Applied Psychological Measurement, 1(3), 385–401.

Roy, Shalini, Melissa Hidrobo, John Hoddinott, Akhter Ahmed. 2019. Transfers, Behavior Change Communication, and Intimate Partner Violence: Postprogram Evidence from Rural Bangladesh. The Review of Economics and Statistics. 101 (5): 865–877. doi: https://doi.org/10.1162/rest_a_00791

Schwarzenberg, Sarah Jane, Michael K. Georgieff, COMMITTEE ON NUTRITION, Stephen Daniels, Mark Corkins, Neville H. Golden, Jae H. Kim, C. Wesley Lindsey, Sheela N. Magge; Advocacy for Improving Nutrition in the First 1000 Days to Support Childhood Development and Adult Health. Pediatrics February 2018; 141 (2): e20173716. 10.1542/peds.2017-3716.

Torrente, C., Aber, J.L., & Shivshaker, A. (2011). Opportunities for Equitable Access to Quality Basic Education (OPEQ): Baseline Report: Results from the Early Grade Reading Assessment, the Early Grade Math Assessment, and children's demographic data in Katanga Province, DRC. New York: New York University

Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, Pollitt E, Carter JA; International Child Development Steering Group. 2007. Child development: risk factors for adverse outcomes in developing countries. Lancet. Jan 13;369(9556):145-57.

Walque, Damien de, Lia Fernald, Paul Gertler, and Melissa Hidrobo. 2017. "Cash Transfers and Child and Adolescent Development." In *Disease Control Priorities, Third Edition (Volume 8), Child and Adolescent Health and Development*, edited by G. Patton D. Bundy, N. de Silva, S. Horton, D. T. Jamison, 325–42. Washington, D.C.: World Bank. https://doi.org/10.1596/978-1-4648-0423-6_ch23.

Wolf AW, De Andraca I, Lozoff B. 2002. Maternal depression in three Latin American samples. Soc Psychiatry Psychiatr Epidemiol. 2002; 37: 169-176.

Figures

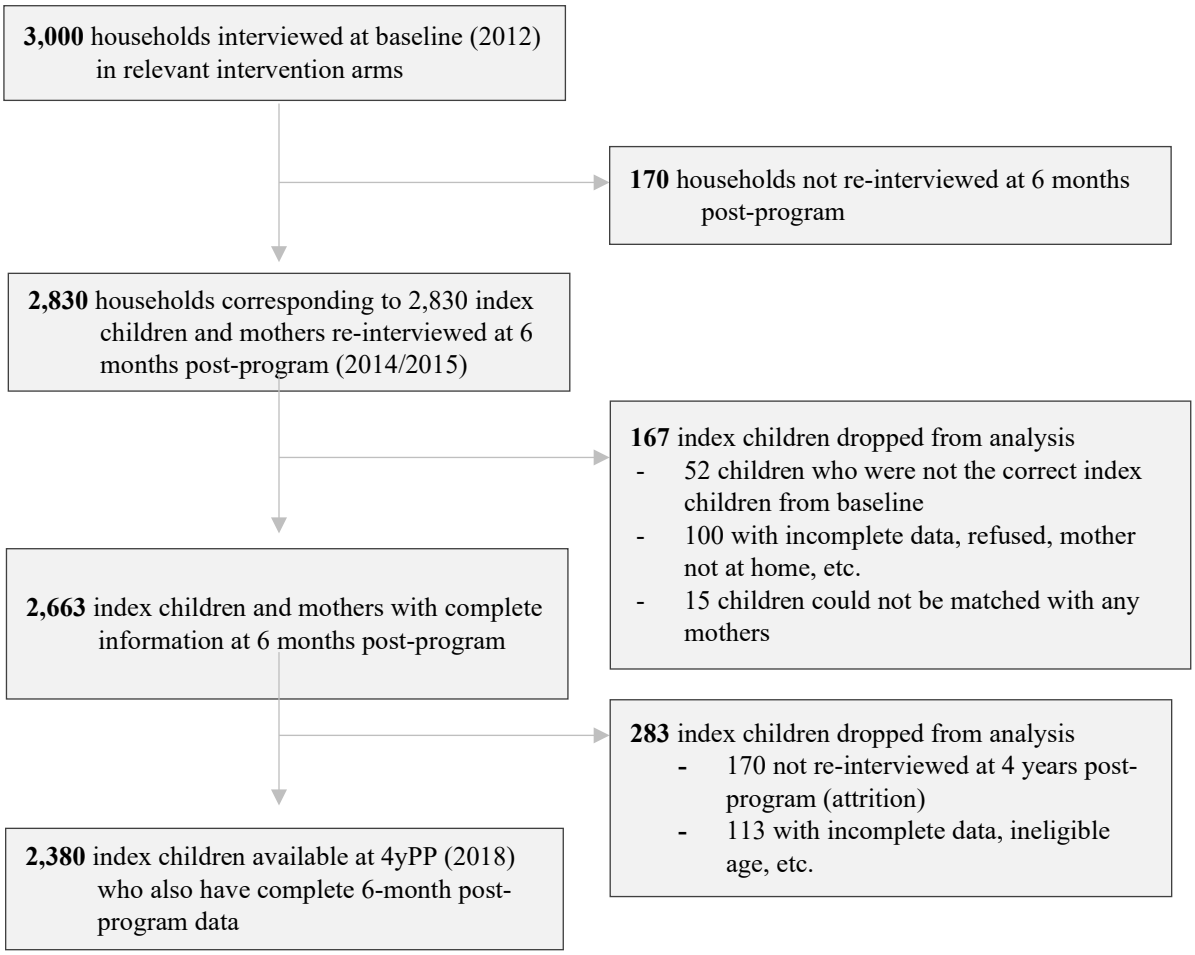


Figure 1: Flow chart of estimation sample

Tables

Table 1a: Impact of treatment arms on maternal outcomes, 6 months post-program, North

	HOME score	Maternal depressive symptoms	Time spent by mother stimulating child (w)	Time spent by all adults stimulating child (w)
Cash	1.32	-0.14	27.06	27.97
	(0.38)***	(1.00)	(43.70)	(49.08)
Cash+BCC	2.79	-1.92	99.98	128.08
	(0.49)***	(0.99)*	(51.01)*	(60.21)**
R^2	0.16	0.04	0.06	0.06
N	1,179	1,179	1,179	1,179
Mean of Control	11.95	9.63	316.89	381.06
P-value: Cash=Cash+BCC	0.00	0.06	0.18	0.11

Control variables include child age and sex, household size at baseline, respondent's age and education level at baseline and total per-capita household monthly expenditure at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. A (w) in front of an outcome means that its values have been winsorized at (1,99) to remove outliers. Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$.

Table 1b: Impact of treatment arms on maternal outcomes, 6 months post-program, South

	HOME score	Maternal depressive symptoms	Time spent by mother stimulating child (w)	Time spent by all adults stimulating child (w)
Food	0.17	0.14	-11.07	-6.97
	(0.37)	(0.72)	(34.86)	(35.53)
Food+BCC	1.50	-0.95	2.74	17.51
	(0.37)***	(0.67)	(38.00)	(38.05)
R^2	0.17	0.02	0.06	0.08
N	1,201	1,201	1,201	1,201
Mean of Control	13.16	9.18	361.70	425.87
P-value: Food=Food+BCC	0.00	0.17	0.73	0.54

Control variables include child age and sex, household size at baseline, respondent's age and education level at baseline and total per-capita household monthly expenditure at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. A (w) in front of an outcome means that its values have been winsorized at (1,99) to remove outliers. Standard errors clustered at the village level. * p<0.1 ** p<0.05; *** p<0.01.

Table 1c: Impact of treatment on home environment outcomes, 6 months post-program, North, by child sex

	HOME environment score		Maternal depressive symptoms		Mother time stimulating child (w)		Total time stimulating child (w)	
	Girls only	Boys only	Girls only	Boys only	Girls only	Boys only	Girls only	Boys only
Cash	1.66	1.07	0.66	-0.85	69.00	-0.41	83.62	-11.73
	(0.50)***	(0.44)**	(1.04)	(1.26)	(56.66)	(55.35)	(63.80)	(60.66)
Cash+BCC	2.79	2.78	-1.16	-2.53	109.63	93.88	153.40	110.83
	(0.61)***	(0.56)***	(1.07)	(1.19)**	(60.60)*	(62.43)	(72.81)**	(69.10)
R^2	0.17	0.07	0.04	0.02	0.08	0.01	0.09	0.01
N	564	615	564	615	564	615	564	615
Mean of Control	11.92	11.98	9.01	10.17	305.49	326.55	365.65	394.13
P-value: Cash = Cash+BCC	0.05	0.00	0.09	0.13	0.53	0.13	0.36	0.08

A higher score on the maternal depression index denotes higher levels of depression. Time spent stimulating the child is measured in minutes over the last 7 days. A (w) indicates that values have been winsorized at (1, 99) to remove outliers. Control variables include child's age at 6mPP, household size at baseline, mother's age and education level at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed). Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 1d: Impact of treatment on home environment outcomes, 6 months post-program, South, by child sex

	HOME environment score		Maternal depressive symptoms		Mother time stimulating child (w)		Total time stimulating child (w)	
	Girls only	Boys only	Girls only	Boys only	Girls only	Boys only	Girls only	Boys only
Food	0.25	-0.13	-0.83	0.94	21.09	-45.32	28.04	-47.85
	(0.47)	(0.48)	(0.95)	(0.94)	(47.94)	(45.16)	(48.45)	(48.76)
Food+BCC	2.16	0.82	-0.84	-1.08	49.21	-35.50	82.78	-40.21
	(0.48)***	(0.48)*	(0.95)	(0.88)	(45.20)	(50.37)	(44.26)*	(54.89)
R^2	0.24	0.01	0.02	0.01	0.08	0.00	0.10	0.00
N	568	633	568	633	568	633	568	633
Mean of Control	12.99	13.31	9.70	8.69	347.80	374.80	408.95	441.80
P-value: Food = Food+BCC	0.00	0.05	0.99	0.04	0.60	0.84	0.32	0.88

A higher score on the maternal depression index denotes higher levels of depression. Time spent stimulating the child is measured in minutes over the last 7 days. A (w) indicates that values have been winsorized at (1, 99) to remove outliers. Control variables include child's age at 6mPP, household size at baseline, mother's age and education level at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed). Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 2a: Impact of treatment arms on child IQ and motor scores, 6 months post-program, North

	Younger children				Older Children			
	Verbal IQ	Performance IQ	Full-scale IQ	2-item motor assessment battery (MAB)	Verbal IQ	Performance IQ	Full-scale IQ	2-item motor assessment battery (MAB)
Cash	0.08	0.03	0.07	-0.09	-0.08	-0.05	-0.06	-0.02
	(0.09)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.09)	(0.14)
Cash+BCC	0.07	-0.17	-0.03	-0.13	0.23	0.12	0.17	-0.01
	(0.08)	(0.08)**	(0.09)	(0.10)	(0.10)**	(0.10)	(0.10)*	(0.12)
R^2	0.06	0.04	0.06	0.02	0.13	0.07	0.13	0.02
N	735	735	735	627	444	444	444	440
Mean of Control	-0.18	-0.05	-0.15	0.01	-0.10	-0.12	-0.16	-0.03
P-value: Cash=Cash+BCC	0.90	0.04	0.28	0.71	0.00	0.08	0.02	0.93

Control variables include child age and sex, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, total per-capita household monthly expenditure at baseline and a dummy for the child being stunted at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. Outcomes are standardized following Attanasio et al. (2020). The full-scale IQ is the sum of verbal and performance IQ scores. The 2-item MAB score is based on the scores on 'threading beads' and 'catching bean bag' subtests; the 'walking heels raised' subtest was dropped due to low variation. Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$.

Table 2b: Impact of treatment arms on child IQ and motor scores, 6 months post-program, South

	Younger children				Older children			
	Verbal IQ	Performance IQ	Full-scale IQ	2-item motor assessment battery (MAB)	Verbal IQ	Performance IQ	Full-scale IQ	2-item motor assessment battery (MAB)
Food	0.09	0.03	0.07	-0.09	-0.01	0.00	-0.01	-0.04
	(0.13)	(0.10)	(0.11)	(0.10)	(0.14)	(0.12)	(0.13)	(0.12)
Food+BCC	-0.09	-0.09	-0.11	-0.05	0.04	0.05	0.02	-0.24
	(0.13)	(0.09)	(0.11)	(0.10)	(0.13)	(0.12)	(0.13)	(0.11)**
R^2	0.08	0.03	0.07	0.02	0.09	0.03	0.07	0.02
N	745	745	745	654	456	456	456	450
Mean of Control	0.17	0.13	0.19	0.12	0.03	0.09	0.12	0.15
P-value: Food=Food+BCC	0.16	0.20	0.12	0.64	0.73	0.73	0.83	0.08

Control variables include child age and sex, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, total per-capita household monthly expenditure at baseline and a dummy for the child being stunted at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. Outcomes are standardized following Attanasio et al. (2020). The full-scale IQ is the sum of verbal and performance IQ scores. The 2-item MAB score is based on the scores on 'threading beads' and 'catching bean bag' subtests; the 'walking heels raised' subtest was dropped due to low variation. Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$.

Table 2c: Impact of treatment on early childhood development outcomes, 6 months post-program, North, by child sex

	Girls only				Boys only			
	Mean of Control	Cash	Cash+BCC	P-value of Cash= Cash+BCC	Mean of Control	Cash	Cash+BCC	P-value of Cash= Cash+BCC
Panel A: Younger children								
Verbal IQ	-0.10	-0.02	-0.02	0.94	-0.25	0.12	0.12	0.97
		(0.12)	(0.11)			(0.11)	(0.11)	
Performance IQ	-0.06	0.02	-0.24**	0.04**	-0.04	0.03	-0.08	0.36
		(0.13)	(0.11)			(0.12)	(0.12)	
Full-scale IQ	-0.09	-0.01	-0.12	0.37	-0.20	0.10	0.05	0.65
		(0.13)	(0.11)			(0.11)	(0.10)	
2-item motor assessment battery (MAB)	0.07	-0.06	-0.25*	0.20	-0.03	-0.10	0.02	0.42
		(0.15)	(0.13)			(0.14)	(0.13)	
Panel B: Older children								
Verbal IQ	0.07	-0.02	0.20	0.12	-0.24	-0.11	0.29**	0.00***
		(0.13)	(0.15)			(0.13)	(0.13)	
Performance IQ	-0.05	-0.02	0.18	0.18	-0.18	-0.06	0.07	0.37
		(0.15)	(0.16)			(0.12)	(0.14)	
Full-scale IQ	-0.02	-0.01	0.20	0.19	-0.28	-0.08	0.16	0.07*
		(0.14)	(0.16)			(0.11)	(0.13)	
2-item motor assessment battery (MAB)	0.06	-0.22	-0.07	0.41	-0.10	0.16	0.02	0.49
		(0.20)	(0.20)			(0.18)	(0.16)	

All outcomes are standardized following Attanasio et al. (2019). Control variables include child's age at 6mPP, child is male (non-disaggregated sample only), household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 2d: Impact of treatment on early childhood development outcomes, 6 months post-program, South, by child sex

	Girls only				Boys only			
	Mean of Control	Food	Food+BCC	P-value of Food= Food+BCC	Mean of Control	Food	Food+BCC	P-value of Food= Food+BCC
Younger children								
Verbal IQ	0.09	0.30*	0.07	0.17	0.26	-0.07	-0.27*	0.20
		(0.16)	(0.16)			(0.15)	(0.15)	
Performance IQ	0.12	-0.02	-0.09	0.60	0.14	0.07	-0.08	0.21
		(0.15)	(0.14)			(0.11)	(0.11)	
Full-scale IQ	0.14	0.18	-0.01	0.23	0.24	-0.01	-0.22	0.14
		(0.15)	(0.15)			(0.14)	(0.14)	
2-item motor assessment battery (MAB)	0.04	-0.11	0.01	0.38	0.20	-0.09	-0.06	0.87
		(0.14)	(0.14)			(0.14)	(0.13)	
Older children								
Verbal IQ	0.10	0.15	-0.08	0.21	-0.02	-0.15	0.16	0.08*
		(0.17)	(0.18)			(0.17)	(0.17)	
Performance IQ	0.11	0.20	0.16	0.86	0.08	-0.11	-0.01	0.48
		(0.16)	(0.18)			(0.17)	(0.17)	
Full-scale IQ	0.18	0.22	0.05	0.39	0.07	-0.17	0.05	0.21
		(0.15)	(0.20)			(0.17)	(0.18)	
2-item motor assessment battery (MAB)	0.14	0.02	-0.31*	0.06*	0.16	-0.12	-0.17	0.75
		(0.16)	(0.17)			(0.16)	(0.14)	

All outcomes are standardized following Attanasio et al. (2019). Control variables include child's age at 6mPP, child is male (non-disaggregated sample only), household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 3a: Impact of treatment arms on home environment outcomes, 4 years post-program, North

	Maternal depressive symptoms	Mother time stimulating child (w)	Total time stimulating child (w)
Cash	-0.74	20.10	18.18
	(0.37)**	(20.95)	(29.47)
Cash+BCC	-1.11	63.14	55.20
	(0.36)***	(25.45)**	(31.53)*
R^2	0.04	0.15	0.09
N	1,173	1,177	1,177
Mean of Control	5.16	229.37	329.72
P-value: Cash=Cash+BCC	0.29	0.11	0.27

Maternal depression is measured using the PHQ-9 raw score (0-27), with lower scores indicating lower levels of depression. Time spent stimulating the child is measured in minutes over the last 7 days. A (w) indicates that values have been winsorized at (1, 99) to remove outliers. Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed). Standard errors clustered at the village level.* $p<0.1$ ** $p<0.05$; *** $p<0.01$

Table 3b: Impact of treatment arms on home environment outcomes, 4 years post-program, South

	Maternal depressive symptoms	Mother time stimulating child (w)	Total time stimulating child (w)
Food	0.17	16.48	-5.25
	(0.28)	(39.65)	(64.44)
Food+BCC	-0.05	23.49	36.96
	(0.29)	(36.35)	(64.10)
R^2	0.05	0.06	0.03
N	1,194	1,196	1,196
Mean of Control	4.81	327.99	490.15
P-value: Food=Food+BCC	0.41	0.85	0.47

Maternal depression is measured using the PHQ-9 raw score (0-27), with lower scores indicating lower levels of depression. Time spent stimulating the child is measured in minutes over the last 7 days. A (w) indicates that values have been winsorized at (1, 99) to remove outliers. Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed). Standard errors clustered at the village level.* $p<0.1$ ** $p<0.05$; *** $p<0.01$

Table 3c: Impact of treatment on home environment outcomes, 4 years post-program, North, by child sex

	Maternal depressive symptoms		Mother time stimulating child (w)		Total time stimulating child (w)	
	Girls only	Boys only	Girls only	Boys only	Girls only	Boys only
Cash	0.01	-1.36	34.52	10.89	31.56	11.76
	(0.43)	(0.47)***	(29.78)	(24.60)	(42.73)	(37.11)
Cash+BCC	-0.48	-1.64	45.41	82.59	27.09	84.34
	(0.41)	(0.50)***	(30.18)	(32.40)**	(42.22)	(38.88)**
R^2	0.05	0.05	0.14	0.16	0.08	0.11
N	562	611	563	614	563	614
Mean of Control	4.64	5.59	236.13	223.63	348.92	313.45
P-value: Cash = Cash+BCC	0.27	0.52	0.71	0.04	0.91	0.12

Maternal depression is measured using the PHQ-9 raw score (0-27), with lower scores indicating lower levels of depression. Time spent stimulating the child is measured in minutes over the last 7 days. A (w) indicates that values have been winsorized at (1, 99) to remove outliers. Control variables include child's age at 4yPP, household size at baseline, mother's age and education level at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed). Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 3d: Impact of treatment on home environment outcomes, 4 years post-program, South, by child sex

	Maternal depressive symptoms		Mother time stimulating child (w)		Total time stimulating child (w)	
	Girls only	Boys only	Girls only	Boys only	Girls only	Boys only
Food	0.48	-0.05	3.85	33.58	5.53	-6.93
	(0.34)	(0.40)	(58.01)	(42.87)	(88.41)	(71.97)
Food+BCC	0.15	-0.23	-6.16	51.07	2.82	68.63
	(0.36)	(0.39)	(47.37)	(42.45)	(75.33)	(81.60)
R^2	0.03	0.07	0.07	0.07	0.03	0.03
N	562	632	564	632	564	632
Mean of Control	4.53	5.08	351.97	305.41	510.20	471.28
P-value: Food = Food+BCC	0.40	0.63	0.83	0.70	0.97	0.30

Maternal depression is measured using the PHQ-9 raw score (0-27), with lower scores indicating lower levels of depression. Time spent stimulating the child is measured in minutes over the last 7 days. A (w) indicates that values have been winsorized at (1, 99) to remove outliers. Control variables include child's age at 4yPP, household size at baseline, mother's age and education level at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed). Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 4a: Impact of treatment arms on standardized child development outcomes, 4 years post-program, North

	Raven	EGRA: listening comprehension	EGRA: letter name recognition	EGRA: non-word reading	EGRA: reading	EGRA: reading comprehension	EGMA: number identification	EGMA: counting	EGMA: smallest/largest number	EGMA: addition and subtraction	SDQ Score
Cash	0.23	0.03	-0.00	-0.10	-0.01	0.01	0.16	0.14	0.06	0.03	-0.12
	(0.10)**	(0.08)	(0.07)	(0.08)	(0.06)	(0.06)	(0.08)**	(0.08)*	(0.08)	(0.08)	(0.11)
Cash+BCC	0.08	0.02	0.04	0.02	0.07	0.02	0.07	0.08	0.07	-0.09	-0.23
	(0.11)	(0.08)	(0.07)	(0.08)	(0.05)	(0.06)	(0.07)	(0.08)	(0.08)	(0.07)	(0.11)**
R^2	0.02	0.04	0.16	0.08	0.14	0.10	0.13	0.13	0.13	0.07	0.04
N	1,179	1,179	1,179	1,179	1,179	1,179	1,179	1,179	1,179	1,179	1,171
Mean of Control	-0.16	0.02	-0.18	-0.10	-0.25	-0.16	-0.17	-0.11	-0.12	-0.03	0.19
P-value: Cash=Cash+BCC	0.16	0.85	0.53	0.08	0.13	0.85	0.19	0.47	0.91	0.14	0.35

Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. All outcomes are standardized following Attanasio et al. (2020). Higher outcomes denote better scores, except for the Strengths and Difficulties Questionnaire (SDQ) Score, where higher value indicate greater behavioral difficulties. Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 4b: Impact of treatment arms on standardized child development outcomes, 4 years post-program, South

	Raven	EGRA: listening comprehension	EGRA: letter name recognition	EGRA: non-word reading	EGRA: reading	EGRA: reading comprehension	EGMA: number identification	EGMA: counting	EGMA: smallest/largest number	EGMA: addition and subtraction	SDQ Score
Food	0.07	-0.11	-0.11	-0.10	-0.11	-0.15	-0.09	-0.10	-0.15	-0.12	0.11
	(0.08)	(0.09)	(0.10)	(0.09)	(0.10)	(0.09)*	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)
Food+BCC	0.03	-0.10	-0.14	-0.15	-0.06	-0.10	-0.09	-0.06	-0.09	-0.10	0.11
	(0.09)	(0.09)	(0.09)	(0.08)*	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)
R^2	0.03	0.07	0.16	0.11	0.13	0.12	0.15	0.14	0.14	0.10	0.04
N	1,201	1,201	1,201	1,201	1,201	1,201	1,201	1,201	1,201	1,201	1,192
Mean of Control	0.04	0.09	0.28	0.24	0.30	0.26	0.19	0.13	0.18	0.16	-0.17
P-value: Food=Food+BCC	0.65	0.92	0.75	0.60	0.61	0.47	1.00	0.71	0.48	0.79	0.99

Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. All outcomes are standardized following Attanasio et al. (2020). Higher outcomes denote better scores, except for the Strengths and Difficulties Questionnaire (SDQ) Score, where higher value indicate greater behavioral difficulties. Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 4c: Impact of treatment on child development outcomes, 4 years post-program, North, by child sex

	Girls only				Boys only			
	Mean of Control	Cash	Cash+BCC	P-value of Cash= Cash+BCC	Mean of Control	Cash	Cash+BCC	P-value of Cash= Cash+BCC
Raven's Standard Progressive Matrices	-0.18	0.19*	0.06	0.25	-0.14	0.25*	0.08	0.24
		(0.11)	(0.11)			(0.14)	(0.14)	
EGRA: listening comprehension	0.07	-0.03	-0.03	1.00	-0.02	0.06	0.06	0.95
		(0.10)	(0.10)			(0.10)	(0.10)	
EGRA: letter recognition	-0.03	-0.06	0.04	0.25	-0.31	0.03	0.02	0.90
		(0.10)	(0.10)			(0.09)	(0.08)	
EGRA: non-word reading	-0.04	-0.11	0.12	0.02**	-0.15	-0.13	-0.11	0.79
		(0.10)	(0.12)			(0.10)	(0.10)	
EGRA: reading	-0.15	-0.07	0.12	0.03**	-0.33	0.03	0.01	0.85
		(0.09)	(0.09)			(0.07)	(0.06)	
EGRA: reading comprehension	-0.05	-0.07	0.03	0.23	-0.25	0.07	-0.01	0.28
		(0.10)	(0.10)			(0.07)	(0.07)	
EGMA: number identification	-0.06	0.06	-0.01	0.48	-0.26	0.23**	0.12	0.21
		(0.11)	(0.10)			(0.10)	(0.10)	
EGMA: counting	0.06	-0.04	-0.03	0.91	-0.26	0.26***	0.16	0.27
		(0.10)	(0.10)			(0.10)	(0.11)	
EGMA: number discrimination	-0.00	-0.05	-0.06	0.88	-0.22	0.14	0.17	0.74
		(0.09)	(0.09)			(0.10)	(0.11)	
EGMA: addition/subtraction	-0.02	-0.07	-0.10	0.82	-0.04	0.10	-0.10	0.03**
		(0.11)	(0.11)			(0.09)	(0.09)	
SDQ Score	-0.02	-0.12	-0.18	0.56	0.36	-0.12	-0.26*	0.33
		(0.12)	(0.12)			(0.13)	(0.14)	

All outcomes are standardized following Attanasio et al. (2020). Control variables include child's age at 4yPP, child is male (non-disaggregated sample only), household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table 4d: Impact of treatment on child development outcomes, 4 years post-program, South, by child sex

	Girls only				Boys only			
	Mean of Control	Food	Food+BCC	P-value of Food= Food+BCC	Mean of Control	Food	Food+BCC	P-value of Food= Food+BCC
Raven's Standard Progressive Matrices	0.03	0.01	-0.01	0.90	0.05	0.10	0.05	0.64
		(0.09)	(0.11)			(0.11)	(0.10)	
EGRA: listening comprehension	0.14	-0.08	-0.14	0.68	0.05	-0.14	-0.07	0.58
		(0.12)	(0.12)			(0.11)	(0.11)	
EGRA: letter recognition	0.40	-0.05	-0.19*	0.26	0.18	-0.17	-0.10	0.53
		(0.14)	(0.11)			(0.11)	(0.11)	
EGRA: non-word reading	0.32	-0.02	-0.16	0.21	0.15	-0.17	-0.13	0.62
		(0.13)	(0.11)			(0.11)	(0.11)	
EGRA: reading	0.45	-0.07	-0.10	0.80	0.17	-0.15	-0.03	0.29
		(0.14)	(0.13)			(0.11)	(0.12)	
EGRA: reading comprehension	0.40	-0.11	-0.19	0.54	0.12	-0.20*	-0.01	0.07*
		(0.14)	(0.13)			(0.10)	(0.11)	
EGMA: number identification	0.26	-0.06	-0.11	0.64	0.13	-0.13	-0.09	0.75
		(0.11)	(0.09)			(0.11)	(0.11)	
EGMA: counting	0.21	-0.07	-0.10	0.78	0.06	-0.11	-0.03	0.51
		(0.11)	(0.10)			(0.11)	(0.11)	
EGMA: number discrimination	0.19	-0.06	-0.13	0.52	0.17	-0.22**	-0.05	0.15
		(0.13)	(0.11)			(0.11)	(0.10)	
EGMA: addition/subtraction	0.13	-0.06	-0.10	0.69	0.19	-0.18	-0.10	0.52
		(0.13)	(0.11)			(0.12)	(0.11)	
SDQ Score	-0.26	0.11	0.00	0.32	-0.08	0.11	0.20*	0.35
		(0.13)	(0.13)			(0.11)	(0.12)	

All outcomes are standardized following Attanasio et al. (2020). Control variables include child's age at 4yPP, child is male (non-disaggregated sample only), household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Appendix Figures

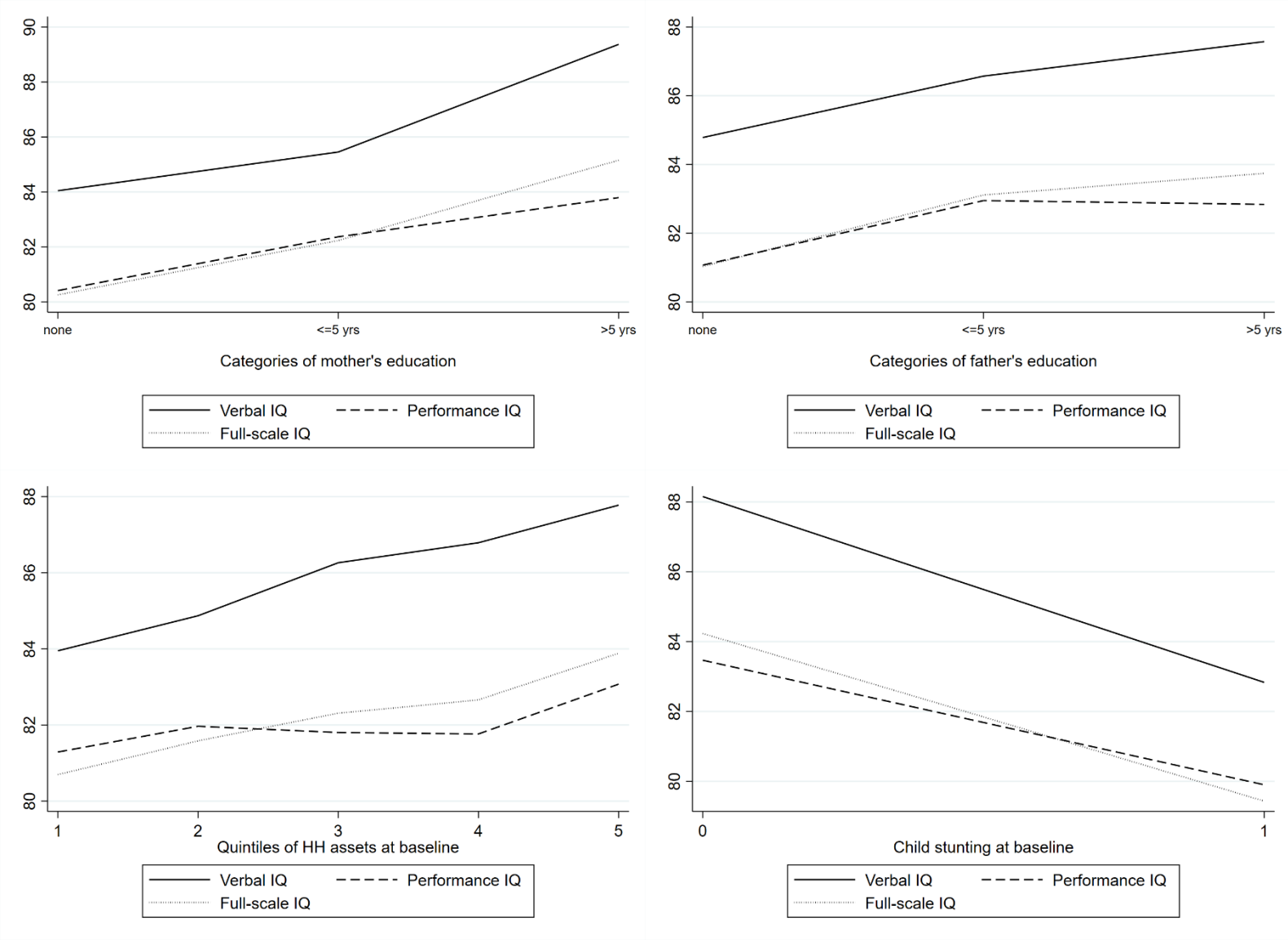


Figure A.1: Association of child IQ scores with select baseline characteristics

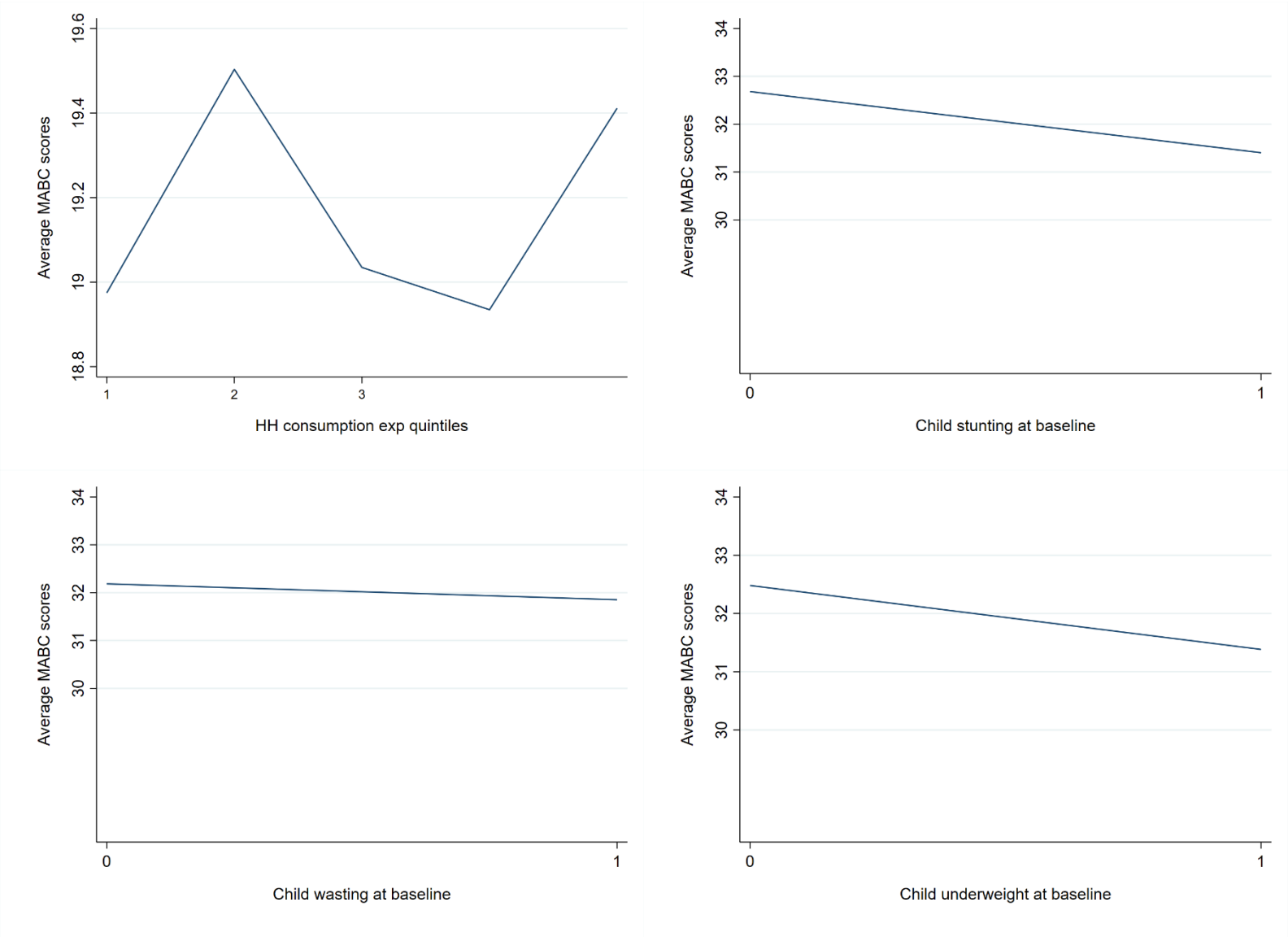


Figure A.2: Association of child motor scores with select baseline characteristics

Appendix Tables

A.1 Attrition and balance tables

Table A.1a: Attrition from baseline sample to analysis sample

	North		South	
	(1)	(2)	(3)	(4)
Pooled treatment	-0.01		-0.01	
	(0.03)		(0.03)	
Cash/Food		-0.03		0.02
		(0.03)		(0.03)
Cash+BCC/Food+BCC		-0.00		-0.03
		(0.03)		(0.03)
Constant	0.22	0.22	0.20	0.20
	(0.02)***	(0.02)***	(0.02)***	(0.02)***
R^2	0.00	0.00	0.00	0.00
N	1,500	1,500	1,500	1,500
P-value: Cash/Food = Cash+BCC/Food+BCC		0.47		0.10

Baseline N=3,000. Non-attrited 6mPP/4yPP estimation sample: N=2,380 (North = 1,179; South = 1,201). Standard errors clustered at the village level.* $p<0.1$ ** $p<0.05$; *** $p<0.01$

Table A.1b: Baseline characteristics by intervention arm, North

	N	Means			P-value of diff.		
		Cash	Cash+BCC	Control	Cash - Control	Cash+BCC -Control	Cash - Cash+BCC
Child's age in months	1,179	13.32	13.24	13.21	0.82	0.95	0.85
Child is male	1,179	0.53	0.50	0.54	0.68	0.23	0.35
Child is stunted	1,179	0.47	0.40	0.47	0.96	0.07*	0.09*
Mother's age	1,179	26.21	27.16	26.54	0.51	0.24	0.06*
Mother's years of education	1,179	2.99	2.81	3.17	0.52	0.21	0.51
Father is in the household	1,179	0.94	0.94	0.93	0.65	0.41	0.71
Household size	1,179	4.91	5.01	4.93	0.89	0.49	0.45
Monthly household expenditure per capita (IHS-transformed)	1,179	7.76	7.74	7.77	1.00	0.55	0.54
p-value from joint F-test					0.97	0.48	0.58

P-values are reported from Wald tests on the equality of means of control and each treatment for each variable. Standard errors are clustered at the village level.
*p<0.1 ** p<0.05; *** p<0.01

Table A.1c: Baseline characteristics by intervention arm, South

	N	Means			P-value of diff.		
		Food	Food+BCC	Control	Food - Control	Food+BCC -Control	Food - Food+BCC
Child's age in months	1,201	12.50	13.47	13.44	0.04*	0.96	0.03*
Child is male	1,201	0.55	0.52	0.52	0.35	0.88	0.46
Child is stunted	1,201	0.41	0.39	0.36	0.17	0.31	0.58
Mother’s age	1,201	27.11	26.21	26.92	0.70	0.16	0.09*
Mother’s years of education	1,201	3.51	3.76	4.09	0.04	0.28	0.37
Father is in the household	1,201	0.87	0.88	0.88	0.80	0.94	0.75
Household size	1,201	5.43	5.26	5.50	0.60	0.11	0.27
Monthly household expenditure per capita (HIS-transformed)	1,201	7.92	7.94	7.91	0.79	0.37	0.51
p-value from joint F-test					0.04*	0.24	0.28

P-values are reported from Wald tests on the equality of means of control and each treatment for each variable. Standard errors are clustered at the village level.
*p<0.1 ** p<0.05; *** p<0.01

Table A.1d: Baseline characteristics by intervention arm, girls only, North

	N	Means			P-value of diff.		
		Cash	Cash+BCC	Control	Cash - Control	Cash+BCC -Control	Cash - Cash+BCC
Child's age in months	564	13.60	13.04	12.82	0.27	0.73	0.37
Child is stunted	564	0.46	0.36	0.45	0.79	0.08*	0.05**
Mother's age	564	26.03	27.20	26.69	0.33	0.48	0.09*
Mother's years of education	564	3.02	2.82	3.54	0.19	0.06**	0.60
Father is in the household	564	0.93	0.96	0.92	0.54	0.04**	0.15
Household size	564	5.04	4.96	4.98	0.72	0.90	0.63
Monthly household expenditure per capita (IHS-transformed)	564	7.79	7.75	7.76	0.60	0.82	0.44
p-value from joint F-test					0.40	0.08*	0.14

P-values are reported from Wald tests on the equality of means of control and each treatment for each variable. Standard errors are clustered at the village level. .* p<0.1 ** p<0.05; *** p<0.01.

Table A.1e: Baseline characteristics by intervention arm, girls only, South

	N	Means			P-value of diff.		
		Food	Food+BCC	Control	Food - Control	Food+BCC -Control	Food - Food+BCC
Child's age in months	568	12.87	13.09	13.09	0.71	1.00	0.71
Child is stunted	568	0.35	0.35	0.29	0.22	0.17	0.91
Mother's age	568	27.16	26.34	26.87	0.65	0.42	0.20
Mother's years of education	568	3.35	3.73	4.06	0.04	0.36	0.27
Father is in the household	568	0.90	0.91	0.88	0.50	0.40	0.86
Household size	568	5.52	5.31	5.49	0.87	0.35	0.29
Monthly household expenditure per capita (IHS-transformed)	568	7.93	7.94	7.92	0.93	0.59	0.65
p-value from joint F-test					0.40	0.40	0.91

P-values are reported from Wald tests on the equality of means of control and each treatment for each variable.
Standard errors are clustered at the village level

Table A.1f: Baseline characteristics by intervention arm, boys only, North

	N	Means			P-value of diff.		
		Cash	Cash+BCC	Control	Cash - Control	Cash+BCC -Control	Cash - Cash+BCC
Child's age in months	615	13.08	13.44	13.54	0.48	0.86	0.57
Child is stunted	615	0.48	0.44	0.49	0.89	0.38	0.49
Mother's age	615	26.38	27.12	26.41	0.95	0.21	0.24
Mother's years of education	615	2.96	2.79	2.85	0.73	0.86	0.62
Father is in the household	615	0.94	0.92	0.94	0.99	0.38	0.39
Household size	615	4.80	5.05	4.88	0.55	0.21	0.12
Monthly household expenditure per capita (IHS-transformed)	615	7.74	7.73	7.77	0.57	0.42	0.80
p-value from joint F-test					0.97	0.66	0.64

P-values are reported from Wald tests on the equality of means of control and each treatment for each variable.
Standard errors are clustered at the village level

Table A.1g: Baseline characteristics by intervention arm, boys only, South

	N	Means			P-value of diff.		
		Food	Food+BCC	Control	Food - Control	Food+BCC -Control	Food - Food+BCC
Child's age in months	633	12.20	13.82	13.77	0.02	0.95	0.02
Child is stunted	633	0.46	0.43	0.42	0.45	0.81	0.59
Mother's age	633	27.07	26.09	26.97	0.88	0.19	0.16
Mother's years of education	633	3.64	3.79	4.12	0.17	0.40	0.67
Father is in the household	633	0.85	0.86	0.89	0.33	0.52	0.78
Household size	633	5.35	5.22	5.51	0.36	0.12	0.48
Monthly household expenditure per capita (IHS-transformed)	633	7.91	7.94	7.90	0.71	0.33	0.57
p-value from joint F-test					0.04	0.42	0.16

P-values are reported from Wald tests on the equality of means of control and each treatment for each variable.
Standard errors are clustered at the village level

A.2 Robustness: 6-month post-program raw composite scores

Table A.2a: Impact of treatment arms on raw IQ and motor scores, 6 months post-program, North

	Younger children				Older Children			
	Verbal IQ	Performance IQ	Full-scale IQ	2-item motor assessment battery (MAB)	Verbal IQ	Performance IQ	Full-scale IQ	2-item motor assessment battery (MAB)
Cash	1.04	0.32	0.77	-0.53	-0.82	-0.55	-0.47	-0.14
	(1.03)	(0.87)	(0.95)	(0.58)	(1.00)	(1.09)	(0.90)	(0.56)
Cash+BCC	0.82	-1.54	-0.28	-0.73	2.45	1.40	1.74	0.03
	(0.96)	(0.74)**	(0.83)	(0.52)	(1.06)**	(1.13)	(0.93)*	(0.49)
R^2	0.06	0.05	0.07	0.02	0.14	0.07	0.13	0.02
N	735	735	735	627	444	444	444	440
Mean of Control	86.83	83.73	83.41	20.08	79.37	76.51	75.63	17.83
P-value: Cash=Cash+BCC	0.83	0.03	0.24	0.71	0.00	0.08	0.02	0.76

Control variables include child age and sex, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, total per-capita household monthly expenditure at baseline and a dummy for the child being stunted at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. Outcomes have not been standardized. Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$.

Table A.2b: Impact of treatment arms on raw composite IQ and motor outcomes, 6 months post-program, South

	Younger children				Older children			
	Verbal IQ	Performance IQ	Full-scale IQ	2-item motor assessment battery (MAB)	Verbal IQ	Performance IQ	Full-scale IQ	2-item motor assessment battery (MAB)
Food	1.09	0.31	0.77	-0.52	-0.06	0.15	0.03	-0.11
	(1.52)	(0.87)	(1.14)	(0.53)	(1.40)	(1.31)	(1.22)	(0.46)
Food+BCC	-1.18	-0.71	-1.12	-0.30	0.49	0.68	0.47	-0.93
	(1.52)	(0.79)	(1.10)	(0.51)	(1.31)	(1.28)	(1.23)	(0.44)**
R^2	0.10	0.07	0.11	0.02	0.09	0.03	0.07	0.02
N	745	745	745	654	456	456	456	450
Mean of Control	91.03	85.36	86.82	20.66	80.84	78.91	78.35	18.57
P-value: Food=Food+BCC	0.14	0.21	0.12	0.69	0.69	0.70	0.74	0.07

Control variables include child age and sex, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, total per-capita household monthly expenditure at baseline and a dummy for the child being stunted at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. Outcomes have not been standardized. Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$.

A.3 Details: 6-month post-program composite score components

Table A.3a: Impact of treatment arms on younger child item-wise IQ and motor scores, 6 months post-program, North

	Verbal IQ		Performance IQ		Motor assessment battery	
	Receptive vocabulary	Information	Block design	Object assembly	Threading beads	Catching beanbag
Cash	0.08	0.03	0.03	0.03	0.02	-0.10
	(0.08)	(0.10)	(0.10)	(0.09)	(0.10)	(0.12)
Cash+BCC	0.11	-0.00	-0.16	-0.09	-0.11	-0.08
	(0.08)	(0.10)	(0.08)*	(0.08)	(0.11)	(0.11)
R^2	0.04	0.05	0.03	0.02	0.03	0.01
N	735	735	735	735	640	637
Mean of Control	-0.20	-0.09	-0.05	-0.06	0.02	-0.00
P-value: Cash=Cash+BCC	0.77	0.69	0.06	0.18	0.17	0.89

Control variables include child age and sex, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, total per-capita household monthly expenditure at baseline and a dummy for the child being stunted at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. Outcomes are standardized following Attanasio et al. (2020). Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$.

Table A.3b: Impact of treatment arms on younger child item-wise IQ and motor scores, 6 months post-program, South

	Verbal IQ		Performance IQ		Motor assessment battery	
	Receptive vocabulary	Information	Block design	Object assembly	Threading beads	Catching beanbag
Food	0.14	0.02	-0.01	0.06	-0.11	-0.03
	(0.13)	(0.11)	(0.10)	(0.09)	(0.08)	(0.12)
Food+BCC	-0.12	-0.03	-0.09	-0.04	-0.08	0.01
	(0.13)	(0.11)	(0.10)	(0.09)	(0.08)	(0.12)
R^2	0.05	0.07	0.04	0.02	0.03	0.01
N	745	745	745	745	663	661
Mean of Control	0.17	0.14	0.11	0.10	0.11	0.06
P-value: Food=Food+BCC	0.05	0.64	0.37	0.25	0.75	0.72

Control variables include child age and sex, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, total per-capita household monthly expenditure at baseline and a dummy for the child being stunted at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. Outcomes are standardized following Attanasio et al. (2020). Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$.

Table A.3c: Impact of treatment arms on older child item-wise IQ and motor scores, 6 months post-program, North

	Verbal IQ			Performance IQ			Performance speed IQ	Motor assessment battery	
	Information	Vocabulary	Comprehension	Block design	Matrix reasoning	Picture completion	Coding	Threading beads	Catching beanbag
Cash	-0.06	-0.08	-0.08	-0.08	-0.06	0.02	0.04	0.12	-0.16
	(0.10)	(0.11)	(0.10)	(0.10)	(0.12)	(0.12)	(0.13)	(0.12)	(0.13)
Cash+BCC	0.25	0.13	0.19	0.10	-0.01	0.16	-0.06	0.11	-0.13
	(0.10)***	(0.12)	(0.10)*	(0.10)	(0.12)	(0.13)	(0.12)	(0.13)	(0.12)
R^2	0.14	0.08	0.07	0.06	0.02	0.08	0.07	0.02	0.02
N	444	444	444	444	444	444	444	444	440
Mean of Control	-0.09	-0.04	-0.13	-0.03	-0.13	-0.06	-0.27	-0.08	0.05
P-value: Cash=Cash+BCC	0.00	0.07	0.02	0.09	0.63	0.21	0.45	0.92	0.81

Control variables include child age and sex, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, total per-capita household monthly expenditure at baseline and a dummy for the child being stunted at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. Outcomes are standardized following Attanasio et al. (2020). Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$

Table A.3d: Impact of treatment arms on older child item-wise IQ and motor scores, 6 months post-program, South

	Verbal IQ			Performance IQ			Performance speed IQ	Motor assessment battery	
	Information	Vocabulary	Comprehension	Block design	Matrix reasoning	Picture completion	Coding	Threading beads	Catching beanbag
Food	0.09	0.02	-0.12	0.00	0.11	-0.15	0.01	0.03	-0.11
	(0.14)	(0.13)	(0.13)	(0.11)	(0.12)	(0.13)	(0.10)	(0.13)	(0.12)
Food+BCC	0.03	0.04	0.03	0.04	0.11	-0.04	-0.01	-0.17	-0.19
	(0.13)	(0.12)	(0.12)	(0.11)	(0.12)	(0.12)	(0.12)	(0.13)	(0.11)*
R^2	0.07	0.07	0.07	0.04	0.02	0.03	0.05	0.03	0.02
N	456	456	456	456	456	456	456	456	450
Mean of Control	-0.05	0.01	0.13	0.02	0.09	0.06	0.23	0.11	0.14
P-value: Food=Food+BCC	0.67	0.90	0.26	0.73	0.99	0.42	0.91	0.08	0.47

Control variables include child age and sex, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, total per-capita household monthly expenditure at baseline and a dummy for the child being stunted at baseline. Missing values in covariates have been replaced by 0s and flags for missing values added. Outcomes are standardized following Attanasio et al. (2020). Standard errors clustered at the village level. * $p<0.1$ ** $p<0.05$; *** $p<0.01$

A.4 Robustness: 4-year post-program raw child development outcomes

Table A.4a: Impact of treatment arms on raw child development outcomes, 4 years post-program, North

	Raven	EGRA: listening comprehension	EGRA: letter name recognition	EGRA: non-word reading	EGRA: reading	EGRA: reading comprehension	EGMA: number identification	EGMA: counting	EGMA: smallest/largest number	EGMA: addition and subtraction	SDQ Score
Cash	0.50	0.05	-0.00	-0.71	0.02	0.02	1.62	0.51	0.33	0.09	-0.61
	(0.22)**	(0.08)	(1.05)	(0.62)	(0.50)	(0.08)	(0.74)**	(0.30)*	(0.41)	(0.26)	(0.51)
Cash+BCC	0.17	0.03	0.78	0.25	0.70	0.03	0.73	0.28	0.39	-0.28	-1.10
	(0.24)	(0.08)	(1.00)	(0.65)	(0.49)	(0.07)	(0.73)	(0.32)	(0.42)	(0.23)	(0.51)**
R^2	0.03	0.08	0.23	0.12	0.22	0.19	0.26	0.26	0.26	0.20	0.05
N	1,179	1,179	1,179	1,179	1,179	1,179	1,179	1,179	1,179	1,179	1,171
Mean of Control	9.31	2.14	14.41	5.06	3.53	0.47	15.41	9.49	6.72	5.39	14.66
P-value: Cash=Cash+BCC	0.18	0.84	0.46	0.08	0.16	0.85	0.19	0.44	0.88	0.13	0.37

Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Higher outcomes denote better scores, except for the Strengths and Difficulties Questionnaire (SDQ) Score, where higher value indicate greater behavioral difficulties. Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table A.4b: Impact of treatment arms on raw child development outcomes, 4 years post-program, South

	Raven	EGRA: listening comprehension	EGRA: letter name recognition	EGRA: non-word reading	EGRA: reading	EGRA: reading comprehension	EGMA: number identification	EGMA: counting	EGMA: smallest/largest number	EGMA: addition and subtraction	SDQ Score
Food	0.16	-0.12	-1.65	-0.76	-0.94	-0.15	-0.94	-0.39	-0.75	-0.40	0.52
	(0.18)	(0.10)	(1.50)	(0.74)	(0.82)	(0.10)	(0.86)	(0.34)	(0.48)	(0.32)	(0.48)
Food+BCC	0.07	-0.11	-2.15	-1.14	-0.50	-0.09	-0.95	-0.28	-0.42	-0.34	0.52
	(0.19)	(0.10)	(1.39)	(0.67)*	(0.79)	(0.10)	(0.83)	(0.33)	(0.45)	(0.30)	(0.50)
R^2	0.06	0.11	0.22	0.16	0.22	0.19	0.24	0.24	0.23	0.20	0.04
N	1,201	1,201	1,201	1,201	1,201	1,201	1,201	1,201	1,201	1,201	1,192
Mean of Control	9.76	2.23	21.63	7.89	8.38	0.93	19.06	10.48	8.33	6.04	12.97
P-value: Food=Food+BCC	0.65	0.95	0.73	0.58	0.57	0.45	0.99	0.78	0.49	0.85	0.99

Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Higher outcomes denote better scores, except for the Strengths and Difficulties Questionnaire (SDQ) Score, where higher value indicate greater behavioral difficulties. Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

A.5 Robustness: 4-year post-program composite EGRA and EGMA indices

Table A.5a: Impact of treatment arms on composite EGRA/EGMA indices, 4 years post-program, North

	EGRA index	EGMA index
Cash	-0.02	0.11
	(0.06)	(0.08)
Cash+BCC	0.04	0.04
	(0.06)	(0.07)
R^2	0.16	0.15
N	1,179	1,179
Mean of Control	-0.23	-0.16
P-value: Cash=Cash+BCC	0.31	0.29

Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Indices are constructed using standardized and equally weighted subscores, following Kling et al. (2007). Standard errors clustered at the village level.* $p<0.1$ ** $p<0.05$; *** $p<0.01$

Table A.5b: Impact of treatment arms on composite EGRA/EGMA indices, 4 years post-program, South

	EGRA index	EGMA index
Food	-0.14	-0.13
	(0.10)	(0.10)
Food+BCC	-0.13	-0.10
	(0.09)	(0.09)
R^2	0.16	0.17
N	1,201	1,201
Mean of Control	0.22	0.16
P-value: Food=Food+BCC	0.92	0.71

Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Indices are constructed using standardized and equally weighted subscores, following Kling et al. (2007). Standard errors clustered at the village level.* $p<0.1$ ** $p<0.05$; *** $p<0.01$

Table A.5c: Impact of treatment on composite EGRA and EGMA indices, 4 years post-program, North, by child sex

	EGRA index		EGMA index	
	Girls only	Boys only	Girls only	Boys only
Cash	-0.09	0.02	-0.03	0.21
	(0.09)	(0.08)	(0.10)	(0.10)**
Cash+BCC	0.07	-0.00	-0.06	0.10
	(0.09)	(0.07)	(0.09)	(0.10)
R^2	0.17	0.15	0.19	0.14
N	564	615	564	615
Mean of Control	-0.12	-0.33	-0.04	-0.26
P-value: Cash = Cash+BCC	0.06	0.78	0.76	0.17

Control variables include child's age at 4yPP, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Indices are constructed using standardized and equally weighted subscores, following Kling et al. (2007). Standard errors clustered at the village level.* $p<0.1$ ** $p<0.05$; *** $p<0.01$

Table A.5d: Impact of treatment on composite EGRA and EGMA indices, 4 years post-program, South, by child sex

	EGRA index		EGMA index	
	Girls only	Boys only	Girls only	Boys only
Food	-0.08	-0.20	-0.07	-0.18
	(0.14)	(0.11)*	(0.12)	(0.11)
Food+BCC	-0.19	-0.08	-0.13	-0.08
	(0.12)	(0.12)	(0.10)	(0.11)
R^2	0.14	0.18	0.18	0.17
N	568	633	568	633
Mean of Control	0.36	0.10	0.19	0.12
P-value: Food = Food+BCC	0.39	0.29	0.62	0.40

Control variables include child's age at 4yPP, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting. Indices are constructed using standardized and equally weighted subscores, following Kling et al. (2007). Standard errors clustered at the village level.* $p<0.1$ ** $p<0.05$; *** $p<0.01$

A.6 Robustness: 4-year post-program child development outcomes, alternative models

Table A.6a: Impact of treatment arms on raw child development outcomes using alternative models, 4 years post-program, North

Model used	OLS	Ordered probit	Tobit (left-censored)	Tobit (left-censored)	Tobit (left-censored)	Probit	Tobit (left-censored)	Tobit (right-censored)	Tobit (left-censored)	Tobit (right-censored)	OLS
Outcome	Raven	EGRA: listening comprehension	EGRA: letter name recognition	EGRA: non-word reading	EGRA: reading	EGRA: reading comprehension	EGMA: number identification	EGMA: counting	EGMA: smallest/largest number	EGMA: addition and subtraction	SDQ Score
Cash	0.50	0.05	0.13	-1.10	0.22	0.01	1.88	0.62	0.48	0.19	-0.61
	(0.22)**	(0.10)	(1.23)	(1.40)	(1.46)	(0.03)	(0.89)**	(0.37)	(0.51)	(0.31)	(0.51)
Cash+BCC	0.17	0.05	0.71	0.08	1.16	0.02	0.92	0.33	0.57	-0.26	-1.10
	(0.24)	(0.09)	(1.25)	(1.38)	(1.41)	(0.03)	(0.90)	(0.41)	(0.54)	(0.26)	(0.51)**
R^2	0.03										0.05
N	1,179	1,179	1,179	1,179	1,179	1,176	1,179	1,179	1,179	1,179	1,171
Mean of Control	9.31	2.14	14.41	5.06	3.53	0.22	15.41	9.49	6.72	5.39	14.66
P-value: Cash=Cash+BCC	0.18	0.97	0.65	0.37	0.50	0.71	0.24	0.46	0.85	0.13	0.37

Ordered probit results reported for EGRA listening comprehension. Left-censored Tobit results reported for EGRA letter name recognition, non-words reading, reading, and EGMA number identification and smallest/largest number discrimination. Right-censored Tobit results reported for EGMA counting and addition/subtraction. Probit results (marginal effects) reported for EGRA reading comprehension (1 = at least one correct answer). OLS results reported for Raven and SDQ for completeness.
Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting.
Higher outcomes denote better scores, except for the Strengths & Difficulties Questionnaire (SDQ) Score, where higher values show more behavioral difficulties.
Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

Table A.6b: Impact of treatment arms on raw child development outcomes using alternative models, 4 years post-program, South

Model used	OLS	Ordered probit	Tobit (left-censored)	Tobit (left-censored)	Tobit (left-censored)	Probit	Tobit (left-censored)	Tobit (right-censored)	Tobit (left-censored)	Tobit (right-censored)	OLS
Outcome	Raven	EGRA: listening comprehension	EGRA: letter name recognition	EGRA: non-word reading	EGRA: reading	EGRA: reading comprehension	EGMA: number identification	EGMA: counting	EGMA: smallest/largest number	EGMA: addition and subtraction	SDQ Score
Food	0.16	-0.15	-1.79	-1.53	-2.19	-0.04	-0.93	-0.40	-0.91	-0.53	0.52
	(0.18)	(0.11)	(1.80)	(1.30)	(1.77)	(0.04)	(0.96)	(0.52)	(0.57)	(0.40)	(0.48)
Food+BCC	0.07	-0.11	-2.36	-2.08	-1.56	-0.03	-1.05	-0.40	-0.59	-0.45	0.52
	(0.19)	(0.11)	(1.67)	(1.14)*	(1.61)	(0.04)	(0.94)	(0.50)	(0.54)	(0.38)	(0.50)
R^2	0.06										0.04
N	1,201	1,201	1,201	1,201	1,201	1,195	1,201	1,201	1,201	1,201	1,192
Mean of Control	9.76	2.23	21.63	7.89	8.38	0.37	19.06	10.48	8.33	6.04	12.97
P-value: Food=Food+BCC	0.65	0.71	0.75	0.67	0.72	0.75	0.90	1.00	0.60	0.82	0.99

Ordered probit results reported for EGRA listening comprehension. Left-censored Tobit results reported for EGRA letter name recognition, non-words reading, reading, and EGMA number identification and smallest/largest number discrimination. Right-censored Tobit results reported for EGMA counting and addition/subtraction. Probit results (marginal effects) reported for EGRA reading comprehension (1 = at least one correct answer). OLS results reported for Raven and SDQ for completeness.
Control variables include child's age at 4yPP, child is male, household size at baseline, mother's age and education level at baseline, father is in the household at baseline, dummy for missing information on mother's baseline age/education, household's monthly consumption expenditure per capita at baseline (IHS-transformed), child is stunted at baseline, dummy for missing values for child stunting.
Higher outcomes denote better scores, except for the Strengths & Difficulties Questionnaire (SDQ) Score, where higher values show more behavioral difficulties.
Standard errors clustered at the village level.* p<0.1 ** p<0.05; *** p<0.01

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available **here**

They can be downloaded free of charge

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS
1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org