Online Appendix

The Impacts of a Prototypical Home Visiting Program on Child Skills

Jin Zhou, James Heckman, Bei Liu, and Mai Lu

January 2, 2024

Contents

List of Figures

List of Tables

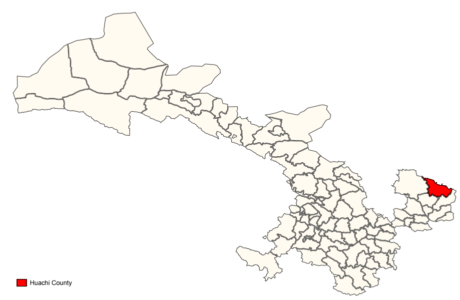
A Conducting the Experiment

Field efforts included recruiting and training interviewers, hiring county-based Ph.D level project directors, and engaging support by the local government and health system. In early 2015, local home visitors were recruited and trained. At the same time, in January 2015, the baseline data were collected, which covered all of the presented households with children under 2 years old in Huachi county. The information includes the Infant-Toddler HOME Inventory for 0–36 months, a household demographic survey, and village-level registration data. Fifty-six villages are randomized into the treatment group, and fifty-five villages are in the control group, for a total of 111 villages in Huachi. Figure A.2 shows the locations of treatment and control group villages in Huachi County.

Figure A.1: The Location of Huachi County

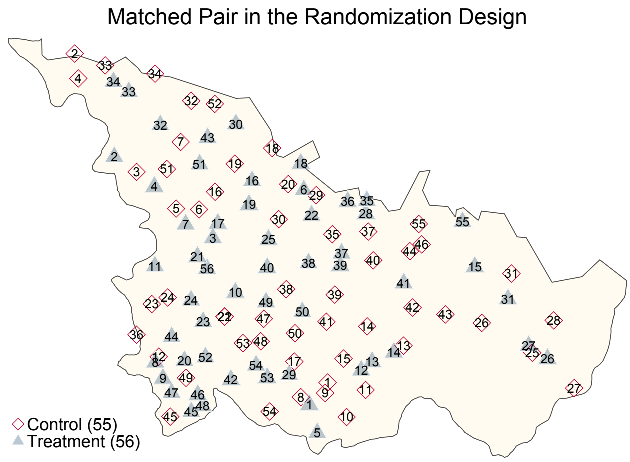
(a) Gansu Province

(b) Huachi County



In January 2016, there was a sample enlargement, including children born between February 9, 2015, and April 30, 2015, for both control and treatment groups. Newly-enrolled children were assessed on their anthropometrics (height, weight, head circumstances, hemoglobin), the Denver II test, the Infant-Toddler HOME Inventory for 0–36 months, and a household demographic survey before the home visiting intervention began.

Figure A.2: The Locations of Randomized Paired Villages in Huachi County



From September 2015 to November 2017, home visitors provided weekly one-hour home visits to the treated children. Home visitors taught caregivers to encourage infants' and toddlers' play-based learning during daily family activities and routine events. Culturally adapted learning materials such as picture books, games, and toys were employed to demonstrate specific interactions.

Between September 2015 and November 2017, two rounds of follow-ups were conducted in July 2016 and July 2017, respectively, which include early childhood development assessments: anthropometrics (height, weight, head circumference, and hemoglobin), the Denver Development Screening Test (2nd ed) for 6–72 months, the Infant-Toddler HOME Inventory for 0–36 months, Early Childhood HOME (36–60 months), and a household demographic survey.

January 2015 Before home visit interventions, baseline information were collected for children with birth dates between April 1, 2013, and November 30, 2014.

September 2015 Initiation of the home visit.

January 2016 Enlarging the enrollment:[1]

- Including 180 children (born between February 9, 2015, and April 30, 2015) for both treatment and control groups.
- Assessed children's health and Denver II scores before home visit intervention.

July 2016 Conducted a family survey and assessed children's health and Denver II test scores for both control and treatment groups.

[1] By January 2016, the data sample included 1566 children: 247 attrited and 76 were newly enrolled. In Section G, we document how we deal with the missing data problem.

July 2017 Conducted a family survey and assessed children's health and Denver II scores for both control and treatment groups.

November 2017 End of home visit intervention.

B China REACH Program Home Visitor Guidelines

B.1 About China REACH

The following appendix is an English translation of the guideline book used to train the home visitors in the China REACH program. It documents how home visitors should approach their work with families and ways to encourage healthy relationships between caregivers and children. It also touches on advising home visitors who can give to caregivers in order to strengthen these attachments.

B.1.1 Intervention Goals

In the course training, home visitors learn the process of child development and how caregivers can promote child development. The home visitors also learn how to conduct home visits and show caregivers how to initiate warm and supportive interactions with their children. Using role-playing and toys, they practice how to work with caregivers to facilitate these interactions.

Curriculum goals for the home visitor:

- Understand the role played by the home visiting program and the home visitor.
- Learn about child development.
- Learn about the skills of conducting home visits and how to introduce new skills and activities to families.
- Learn how to record home visits throughout the course progress.
- Make toys for home visits.

Learning goals for the caregivers in the intervention:

- Acquire knowledge about child development.
- Improve ways of talking, playing games, teaching, and interacting with children.
- Be able to use daily activities and household items to teach and play with children.
- Improve self-confidence and gain happiness in the process of promoting children's growth.

Learning goals for children in the project:

- Improve language ability and advance intellectual development.
- Improve behaviors and develop their social and emotional abilities.

B.1.2 Children's Intellectual Development

With the help of adults, children explore things in their environment, play with and communicate with adults, and imitate their behaviors.

- Children aged 6-24 months

  Adults show children how to interact with the following materials:
  - Items of different colors, shapes, and materials.
  - Items that can be put in or taken out, items that make a sound when colliding, items that can be stacked together, and items that can be opened and closed.
- Children aged 24-48 months

  Children learn from the following activities:
  - Playing games
  - Imitating, pretending to be an adult
  - Playing puzzle games
  - Building objects and matching them. Playing games is very important for children, and it is essential for their intellectual development. When adults are involved and describe what the children are doing, they learn the most and are the most effective.

B.1.3 How to Improve Children's Language Skills

The home visitor aims to help the caregiver support the child's language development by facilitating interactions that focus on the following methods:

- Respond to the child's voice, words and questions.
- Introduce new sounds and words.
- Talk to your children as much as possible and describe what you are doing (for example: "I am making breakfast for you now," "I am washing these dirty clothes.")
- Call out the names of objects and people at home and outside.
- Tell the children what they are holding, what they want to reach, and what they are interested in.
- Do a naming game: say the name of the object and let the child point it. Then point to the item and ask the child to call out the item's name.
- Give praise to your child when he/she uses new words correctly.
- Expand the child's vocabulary (e.g., When the child says, "Look, dog," the caregiver says, "Yes, this is a big white dog.")
- Watch albums and photos with them.
- Use situations in daily life (such as bathing, coaxing him/her to sleep, eating, etc.) to speak new words and words learned before.
- Play role-playing games with your child (for example, say "Do you like your baby? Let's feed her something.")

B.1.4 How to Help Children Develop Their Social and Emotional Abilities

The child needs a caregiver that is:

- Someone they trust (they respond to their needs, such as hungry, unhappiness, etc.).
- Someone who expresses love to them (such as hugging, comforting, kissing them, speaking to them softly and gently, telling them that they love them).
- Someone who communicates with them (responds to their voice).
- Someone who understands them (knows what they like to do, what makes them happy/unhappy), and plays with them.
- Someone who is always by their side (the caregiver they have always been familiar with).

This will make them feel safe, confident, and happy. They will develop a strong, secure attachment to this person, and the strength of this attachment will affect the way they get along with other people in the future and their happiness in later life.

B.1.5 How to Improve Children's Self-Confidence

Children living in poverty usually lack self-confidence. The following activities can help them strengthen their self-confidence.

- Refer to the children using their names as much as possible.
- Praise the children on their personal images/behaviors (For example, "You are already a big kid," "You are so smart," "You are a helpful kid").
- Discuss with them on what they are doing ("I see you flipping through the picture album and eating," "I see you like this doll/that toy").
- Always listen and respond to the child's voice.
- Allow him/her to make their own choices.
- Give them praise when they successfully complete tasks or play games well.
- Make sure that the children experience successes more often instead of failures.
- Ask the caregiver to make a toy bag for the child, put it somewhere in the house, and then tell the kid that the bag belongs to him/her.

What will hurt children's self-confidence:

- Regular punishments, especially beating, ridiculing, and blaming.
- Frequent failure experience and/or often giving the children difficult tasks.

B.1.6 Home Visit Guide

The home visitor helps the caregiver introduce new activities to the child during visits. Activities such as making toys, completing puzzles, and making picture albums are part of what the home visitor brings to the visit with each family. The goal is to have the caregiver interact warmly, and spontaneously with the child. A home visitor is assigned to each family at enrollment and she continues to see the family for the duration of the program. This helps build a relationship with the family, which in turn helps form trust.

- It is very important not to change the home visitor of each family in the project.
- It is also very important to complete each home visit. Higher the frequency of home visits, better the child's growth.

B.1.7 Caregiver's Role

Caregivers play a crucial role in their children's developmental success. When a caregiver learns how to support his or her growing child, that knowledge can be beneficial long after the intervention is over.

Home visitors lead home visits, but caregivers and home visitors share responsibilities. Caregivers eventually take over the duties of home visitors.

We need to encourage caregivers to decorate the home into a more stimulating environment, play with children with home and outside objects, talk to and play with children in daily activities, teach them new knowledge, and spend time with them to play with toys, look at the picture album and communicate with words.

B.2 Training Sessions

B.2.1 Home Visit

- How does the home visitor establish this positive relationship?
  - Sits upright
  - Asks the caregiver what activities and games she has done with her child recently
  - Actively listens (recognizes and thinks about what the caregiver said)
  - Praises her
- How do you tell if the caregiver is satisfied with this home visit?
  If the caregiver
  - Smiles or laughs
  - Is able to complete the activity well
  - Plays with kids with pleasure
- What does the home visitor do to make the caregiver satisfied with the home visit?
  - Praises the caregiver
  - Praises the child
  - Makes sure the caregiver and the child know they are doing well

B.2.2 How to Help Children Learn

- Background information
  - Children learn knowledge and skills through imitating, exploring and experimenting.
  - Children learn through multiple repetitions.
  - Children learn the best when adults are involved.
  - When introducing a new activity, find out the most accessible part for the child.
- How to help children learn
  - Give the child new materials and allow him enough time to explore the materials independently. Observe what he did and praise it (for example, "Wow, you picked up a building block. Look, he picked up a building block, so smart.").
  - Explain the goal of the new activity to the caregiver (e.g., put the blocks into and out of the container).
  - Demonstrate new activities to children.
  - Encourage children to do activities and allow them to practice independently. Praise the child when she tries to perform activities.
  - Involve mothers in children's activities.
  - Whenever the caregiver tries to participate in activities, remember to praise her.
  - Let children practice more independently.
  - When mothers and children participate in activities, praise them.

B.2.3 How to Teach Mothers

- How do you tell if the caregiver has understood how to do this activity with the child?
  The caregiver:
  - Is able to do this activity with children
  - Praises the child
- How does the caregiver know what she should do?
  The home visitor:
- - Explains clearly
  - Demonstrates activities
- How do you judge if mothers and children like this activity?
  The caregiver and the child:
  - Are happy
  - Can perform this activity well
- What makes them like this activity?

  - The home visitor praises caregivers and children, demonstrates activities, and gives caregivers time to practice.

B.2.4 The Importance of Praise/How to Make Home Visits Interesting

- Praise can help home visitor to establish a positive relationship with the caregivers and help them build self-confidence.
- Praise can also improve children's language skills, promote the development of social and emotional skills and self-confidence.
- Praising a child can make her feel smart and willing to do this activity.
- Compliment your child in the following situations:
  - When playing with toys
  - When communicating with movement, voice, or speech
  - When having fun during a home visit
  - When trying an activity, even when it is unsuccessful
  - When performing activities
- Praise can be verbal or non-verbal. Non-verbal praise includes applause, high five, hug, smile, etc.
- Importance of increasing the fun of home visits
  - Better establish a relationship with caregivers and children
  - Enhance the confidence of caregivers and children
  - Make caregivers feel more comfortable and natural
  - Let children feel at ease, confident and happy
  - Motivate children to learn
  - Increase children's participation

B.2.5 Listen, Understand, and Respond to Children

- Young children communicate with others by pointing, making noise, reaching out, crying, and smiling. As children grow older, they begin to speak a few words and then a few sentences.
- Explain to the caregiver that they should always tell their children what they are doing and what they want to express.

B.2.6 Build a Good Relationship with Caregivers

- The importance of establishing a good relationship with the caregiver. Good relationships

- Encourage her to participate in activities
- Build caregivers' confidence
- Make caregivers more willing to listen to the advice of a parenting counselor
- Make caregivers more willing to share personal difficulties and successes

- Home visitors and caregivers establish a good relationship through the following methods:
  - Sit upright
  - Ask about activities that have been done with the child
  - Actively listen (recognize and think about what mom said)
  - Praise her

B.2.7 Understand Difficulty

- It is very important to give the child activities that match his ability. If it is too easy, the child will feel bored. If it is too difficult, the child will feel frustrated and shocked.
- If the task is difficult to complete, you can break it down into simple steps or reduce the difficulty.
- If your child can complete simpler tasks, try to increase the difficulty.
- It is important to ensure that children understand and are competent in simple activities when they start higher-level games.
- When the child himself can repeat an activity correctly without the help of others, we can make it more difficult for him.

B.2.8 Give Feedback to the Caregiver

- When caregivers are involved in a certain activity or perform well, the home visitor should explain what they are doing well.
- Feedback should also make caregivers feel more comfortable during home visits and give them enough confidence to participate in activities on their own.
- Always praise the caregivers. You can praise them in the following situations: when they
  - Participate in activities with their children
  - Praise their children
  - Chat with their children
  - Tell you what their children did/can do
  - Tell you what activities they and their children did (example)
- You can praise the caregiver in these ways:

  - Recognize what the caregiver has done
  - Smile at the caregiver
  - Use specific praise to let the caregivers know what they are doing well

B.2.9 Use Daily Activities to Help Children Learn

Children learn by observing and imitating others. They will eventually want to do things on their own and explore their surroundings. Caregivers can teach them new things through daily activities, talking, and playing with the children. Caregivers should be encouraged to use things at home and outside for children to play.

B.2.10 First Home Visit

The first home visit is especially important because parents must understand the project's goals, the content of the activities, and what they are going to do.

- The home visitor starts the home visit by greeting the caregiver and introducing herself. They should both understand how to address each other.
- The home visitor asks the caregiver about the child:
  - The home visitor explains the project
  - The home visitor asks about family information and who lives at home. She also notes whether there are other children in the family and their ages.
  - Before the end of the visit, the home visitor should make an appointment with the caregiver for the next visit
- Explain the project
  - Number of home visits
  - What you will do and its importance
  - Impacts on child development
- Ask about family information and who lives at home. Note whether there are other children in the family and their age.
- Make an appointment with your caregiver for the next home visit.
  - Arrange a suitable time for you, the caregiver, and the child.
  - Record the phone number or other contact information.

B.2.11 Involving Other Family Members

Family members other than the primary caregiver in the program are very important in a child's life.

- Some fathers and grandparents showed great interest in the intervention, and their support was very helpful. Including these family members will contribute positively to the child's experience.
- Include any adults living in the family, such as fathers or grandparents, who are willing to take care of their children and participate in activities.
- Bring toys (picture albums/crayons and paintbrushes) to other children living in the home for their play. Some activities, such as games, are more appropriate for other children to include.
- Why is it important for other family members to include?

  When including family members, children will:
  - Feel happy
  - Have more fun
  - Build confidence

B.2.12 Promote Positive Behavior

Children learn by observing and imitating others. They will eventually want to do things themselves and explore their environment. Good management of children's behavior will make the activity smoother and make everyone feel more comfortable.

Here are several ways to manage your child's behavior. We can play with children, praise them, give them choices, keep them safe and distract them.

We can praise children when they:

- Do something well
- Try to do something (even if they do it wrong)
- Show the real themselves

Giving children the right to choose makes them feel that they can control what happens. For example, allow children to choose which book to read or which food to eat.

As children grow up and become more independent, they will want to explore more. Avoid always saying "no" to your child. One way is to make the home environment safer for children to explore. Put away what the child can reach and break.

Distract children and keep them away from things they cannot touch or things they cannot do instead of saying "no."

B.3 Curriculum

The development of skills in young children has been extensively studied and theorized over the years (e.g., Uzgiris and Hunt (1975) and Palmer (1971) are major references). The China

REACH program curriculum is adapted from the Jamaican Reach Up and Learn program, which is designed to focus on a child's ability to complete sequences of tasks ordered by progressing difficulty levels based on general child development patterns. In general, children's skill development depends on a number of factors such as caregiver involvement, cultural environment, nutrition, child endowment, etc. To better understand how the skills develop over time, it is necessary to analyze the measures used to evaluate children's multidimensional skills. Based on the main content of tasks, the tasks in the curriculum cover four domains of skills.[2] The categories help researchers understand how the main types of skills developed based on the measures in the curriculum. Next, we document all the tasks in the China REACH curriculum by four domains of skill types: fine motor, gross motor, language, and cognitive skills.

B.3.1 Skills Taught in the Curriculum

Fine motor, gross motor, language, and cognitive skills are taught. Within each skill group, skills are ordered by difficulty level following the patterns developed by Palmer (1971). For example, there are seven difficulty levels for fine motor drawing lessons for.[3,4] In general, higher difficulty level of skills includes new content. For example, difficulty level 2 is to mimic circles. The skills at difficulty level 3 include drawing straight lines. We document how tasks in different difficulty levels are categorized.

Using Fine Motor Drawing lessons to explain in details: the lessons focus on a child's ability to use writing utensils with increasing skills. First, a child is asked to hold the utensil to make markings. Next, the child incorporate more and more cognitive skills to complete the tasks. They then begin by copying markings made by an adult. As skill levels progress, they are asked to make the marking after only a verbal command from the adults. Finally, the child progresses from abstract shapes to representative drawings (See Table B.1.).

[2] We are aware that skills do not develop in isolation, fine motor skills require cognitive input and language skills develop in tandem with gross motor functions.

[3] The standard of generating the difficulty levels are based on the understanding of the content in the skills.

[4] The difficulty level in our content only has ordinal meaning, not cardinal meaning.

Table B.1: Skill Levels for Fine Motor (Drawing) Lessons

Difficulty Level	**Task Content**
Level 1	Doodle using crayons
Level 2	Mimic draw circles
Level 3	Mimic circles and draw straight lines
Level 4	Draw a circle, vertical line, and horizontal line
Level 5	Draw circles, many lines, and crossed lines
Level 6	Draw a cross (or T), curves, and zigzag curves
Level 7	Draw caterpillars

In addition to tasks of different difficulty levels, the curriculum features multiple lessons and assessments at the same difficulty level l. The description of difficulty level categories is listed in this section. For example, there are six assessments at difficulty level 3 for fine motor drawing skills and only two assessments at difficulty level 2.

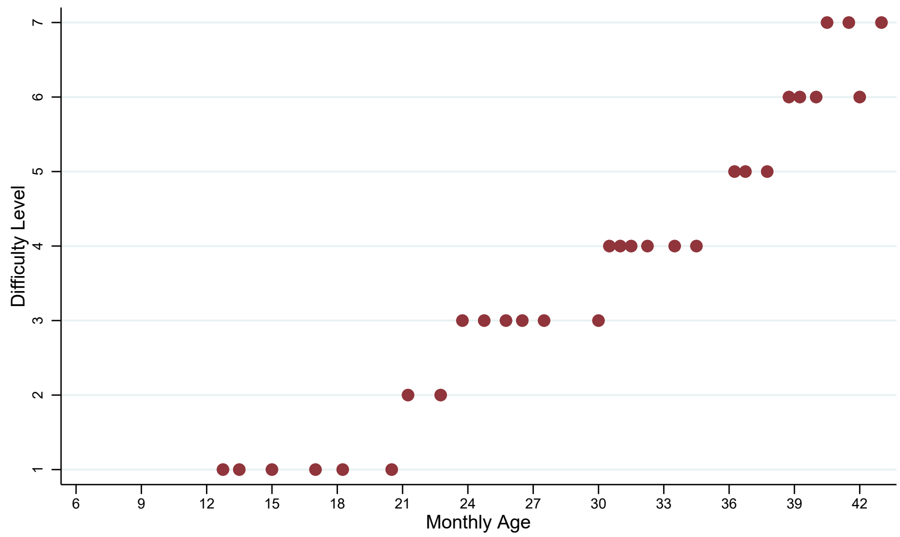
Figure B.1 shows the timing of each fine motor drawing assessment in the curriculum design. For the designated skills, difficulty level 1 covers from 12 months and three weeks to 20 months and two weeks. This timing means that when the child is 12 months and three weeks old, the home visitor will teach her the first fine motor drawing skill. When she is 20 months and two weeks old, the home visitor will teach her the sixth lesson at difficulty level 1. In general, higher difficulty levels appear at later weekly ages. However, there can be some overlap across difficulty levels. For example, in Figure 2, by the time difficulty level 7 of fine motor lessons start, the last lesson of level 6 remains unfinished. In Figure B.1, when fine motor lessons at difficulty level 7 start, the student still receives lessons at difficulty level 6. Circling back is a strategy designed to solidify a child's understanding of a concept.

Another example concerns cognitive skill categories. Cognitive skills have different dimensions. In the curriculum, the cognitive skills taught cover spatial, knowledge of objects and object functions, order and number, etc. Using knowing objects and object functions as an example: cognitive skill difficulty levels are defined based on the abstract concepts shown in Table B.2, such as the child's proficiency in understanding the objects. Seventy-four lessons are sorted into the listed 13 ordered difficulty levels.[5] It covers the process of how the child learns to know an object and understand the function of the object.

The lessons in the cognitive knowledge of objects unit progress from a simple understanding of the concept of pictures by acknowledging with vocalizations, to using receptive (heard) language to identify certain pictures. Receptive language is a skill developed prior to the expressive language where a child forms words to communicate. The child must use his

[5] The difficulty level in our content only has ordinal meaning, not cardinal meaning.

Figure B.1: The Timing of Fine Motor Skill (Drawing) Tasks across Difficulty Levels



Note: Level 1: Doodle using crayons; Level 7: Draw caterpillars.

or her expressive language to complete the following lessons, which increase with difficulty as they must develop more and more language to identify an increasing number of images. To progress through level 7 and beyond, the child must display an increasingly sophisticated understanding of the stories presented, first simply naming actions, then answering questions, then talking abstractly about a story. Levels 10, 11, 12, and 13 ask the child to take the information presented and build on it by discussing the uses of objects presented and making connections with other images.

Figure B.2 shows the timing of each cognitive (knowing objects and understanding the object's function) level in the curriculum. According to the curriculum content, the number of lessons varies across difficulty levels. Table B.3 presents detailed information about the six lessons (and assessments) that are labeled as difficulty level 1 directed to ten-month to 15 month-old curriculum content. In Table B.3, all lessons relate to the activity of looking at the pictures or objects and vocalizing, which does not require the child to name or identify the object.

Table B.2: Difficulty Level List for the Cognitive Understanding Objects Lessons

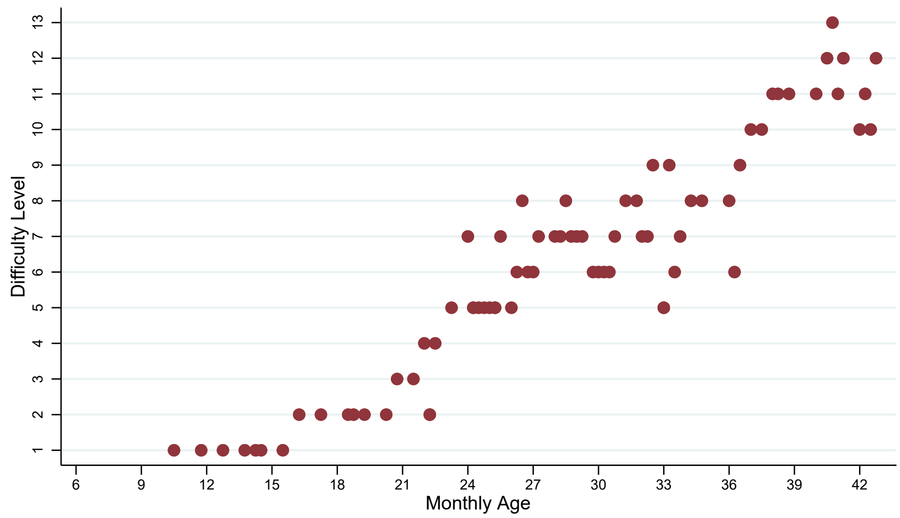
Level 1	The child can look at the pictures and vocalize
Level 2	Name the objects and ask the child to point to the corresponding pictures
Level 3	The child can name the objects in one picture, and point to the named picture
Level 4	The child can name the objects in two or more pictures, and point to the named picture
Level 5	The child can point out named pictures, and say names of three or more
Level 6	The child can point out the picture mentioned and correctly name the name of six or more pictures
Level 7	The child can talk about the pictures, answer questions, understand, or name the verbs (eat, play, etc.)
Level 8	The child can follow the storyline, name actions, and answer question
Level 9	The child can understand stories and talk about the content in the pictures
Level 10	The child can keep up with the development of the story
Level 11	The child can say the name of each graphic, discuss the role of each item, and then link the graphics in the card together
Level 12	The child can name the things in the picture, link different pictures together, and discuss some of the activities in the pictures
Level 13	The child can name the things in the picture and talk about the function of objects

Table B.3: Cognitive Skill Task Content: Look at the Pictures and Vocalize (Level 1)

Difficulty Level	Month	Week	Learning Materials	Content
1	10	2	Picture book A	The baby makes sounds when looking at the pictures
1	11	3	Picture book B	The baby looks at the pictures and vocalizes
1	12	3	Picture book A	The child makes sounds looking at the pictures
1	13	3	Picture book B	The child makes sounds looking at the pictures
1	14	1	Picture book A	Mother and child look at the pictures together, and the mother lets the child vocalize and touch the pictures
1	15	2	Picture book B	Mother and child look at the pictures together, and the mother lets the child vocalize and touch the pictures

In sum, the curriculum targets lessons for multiple levels of skill at each weekly age. For each type of skills, the difficulty levels are constructed by the content of the tasks and the guideline of Uzgiris and Hunt (1975) and Palmer (1971). The terms of the number of lessons within each difficulty level varies. We follow these scholars and assume that each level is a quantum of understanding that is comparable across children. We use achievement at each level of skill as our measure of knowledge.

Figure B.2: The Timing of Cognitive Skill (Understanding Objects) Tasks across Difficulty Levels



Note: Level 1: Look at the pictures and vocalize; Level 13: The child can name the things in the picture and talk about the function of objects.

B.3.2 Fine Motor Skill

Fine motor skill involves finger movements, such as grasping, releasing and stitching, and drawing and writing skills. Here we consider two types of fine motor skills: (1) finger movements related to grasping, releasing, stitching; and (2) the movements related to drawing and writing ability. This task evaluates whether a child can grasp the writing instrument and make marks, scribbles, and shapes. It is not writing ability as in letters or words.

The first category is related to finger movements regarding grasping, releasing, stitching.[6] In Table B.4, tasks progress from basic activities like holding and moving an object that require limited precision with the fine muscles of the hands to manipulating the object with movements that need incrementally more dexterity (like rotating the object) to complex tasks requiring finer and finer finger control, like unscrewing the top. Finally, tasks that require the most hand dexterity, as well as hand-eye coordination, come last.

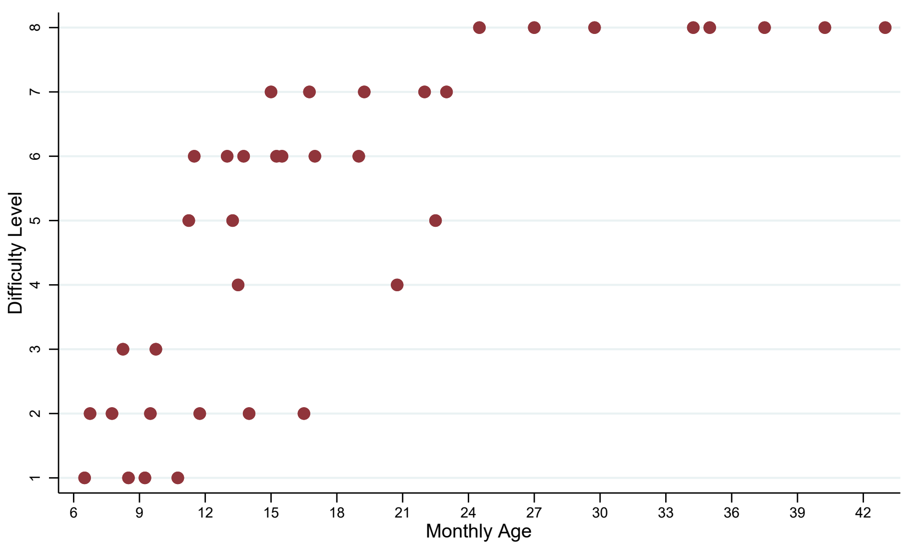
[6]These milestones are justified at `https://www.chrichmond.org/therapy-services/occupational-therapy/developmental-milestones/fine-motor-skills-birth-to-2-years` and `http://www.kamloopschildrenstherapy.org/fine-motor-skills-infant-milestons`.

Table B.4: Difficulty Level List for Finger Movement Tasks

Level 1	Rattle the bottle
Level 2	Shake and beat the drum with two hands
Level 3	Pull strings to get toy
Level 4	Rotate, push
Level 5	Place small objects into the bottle, shake it, and unscrew the lid
Level 6	Put small container into a larger container
Level 7	Take the ring off and slip the ring onto the bottle
Level 8	String beads

Figure B.3 gives the timing of each finger movement tasks in the curriculum.

Figure B.3: The Timing of Fine Motor Skill (Grasping, Releasing Actions) Tasks across Difficulty Levels



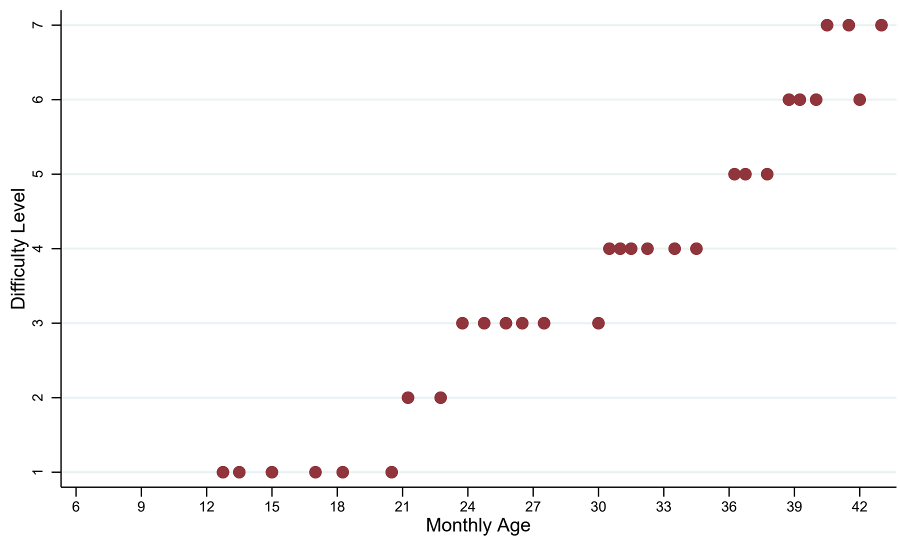
Note: Level 1: Rattle; Level 8: String beads.

The second category is related to drawing and manual writing ability. The fine motor drawing tasks in Table B.5 focus on a child's ability to use a writing tool with increasing skills. First, a child must be able to hold the tool to make markings. Next, the child must incorporate increasingly complex cognitive skills to complete the tasks. They start by imitating markings made by an adult. Then, when skill levels progress, they must make the marking after only a verbal command from the adult. Finally, the child progresses from abstract shapes to representative drawings.

Table B.5: Difficulty Level List for Fine Motor Drawing Tasks

Level 1	Doodle using crayons
Level 2	Mimic draw circles
Level 3	Mimic circles and draw straight lines
Level 4	Draw a circle, vertical line, and horizontal line
Level 5	Draw circles, many lines, and crossed lines
Level 6	Draw a cross (or T), curves, and zigzag curves
Level 7	Draw caterpillars

Figure B.4: The Timing of Fine Motor Skill (Drawing) Tasks across Difficulty Levels



Note: Level 1: Doodle using crayons; Level 7: Draw caterpillars.

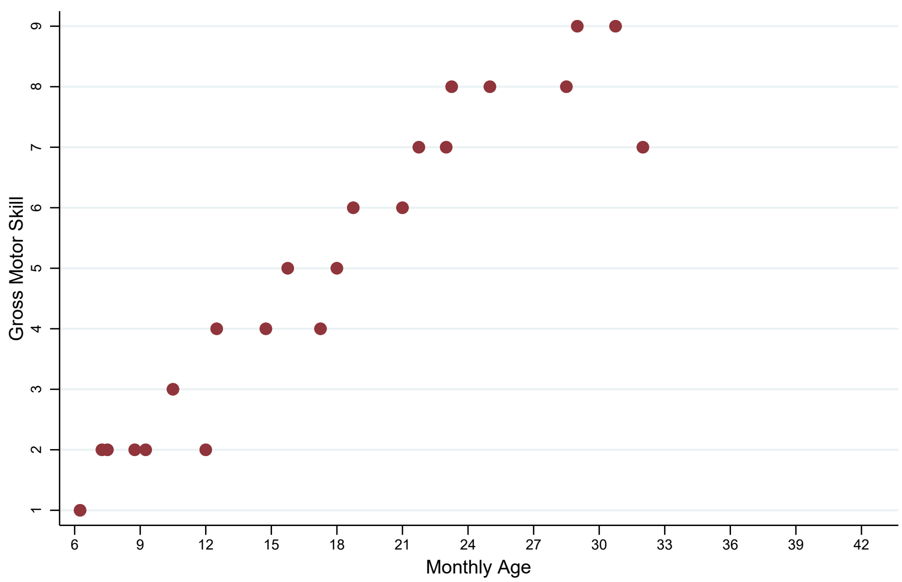
B.3.3 Gross Motor Skill

Gross motor skill is any skill that requires movement and precision of large body muscles. Crawling, creeping, walking, throwing and dancing are all examples of gross motor skills. The designated gross motor tasks start with a relatively simple activity, touching the ball, requiring the child only to move one hand to the object. Next, the child must be able to move his or her entire body to interact with the toy. After mastery over those tasks, the child uses both gross motor skills and newly found cognitive ability to interact with the toy in increasingly complex ways. Pushing a toy requires coordination, standing, and walking skills. However, the child is still using the toy as a walking aid at this point. To progress to the next tasks, not only will the child have to master walking independently, but will also use the toy in a way that suggests intentionality (e.g., pulling, throwing). The final tasks require the child to integrate cognitive knowledge of direction, descriptive words, and gross motor mastery of balance.

Table B.6: Difficulty Level List for Gross Motor Tasks

Level 1	Let the child touch the ball
Level 2	The child moves (crawls) and follows the ball
Level 3	Roll the ball
Level 4	Push the toy when walking
Level 5	Pull the toy
Level 6	Pull and walk forward or backward
Level 7	Throw ball backward, forward, upward and into a target
Level 8	Move forward or backward. Child can understand "upward," "downward," "inside of," "outside of," "stop," "go," "fast," "slow."
Level 9	Hold the soft ball on his or her head stably while walking

Figure B.5: The Timing of Gross Motor Skill Tasks across Difficulty Levels



Note: Level 1: Let the child touch the ball, Level 9: Hold the soft cloth ball on his head stably while walking.

B.3.4 Cognitive Skill

Cognitive skill is broadly defined as a child's ability to apply what they have learned previously for new situations. This skill involves logic, problem-solving ability, memory, attention, and so on.

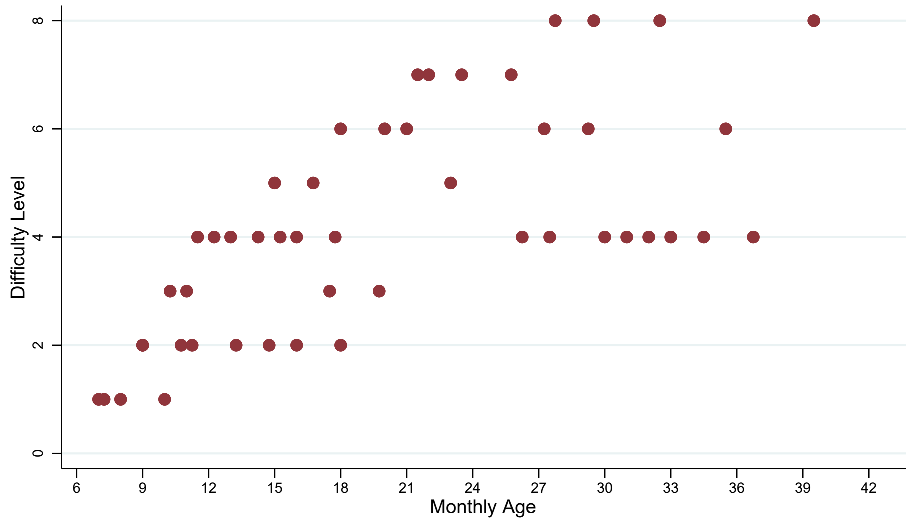
B.3.4.1 Spatial Skill

Spatial skill relies on a child's understanding of the three dimensional world. Comprehending concepts of relative positioning—"inside of," "around," and "next to" are the basics of this skill. The progression of these skills follows the child as he or she learns concepts that are more and more abstract. Beginning with "in" and "out" and progressing to "underneath," "around," "up," "next to," and "close to." As the tasks become more difficult, the child is expected to manipulate objects to demonstrate knowledge and understanding of these concepts.

Table B.7: Difficulty Level List for Cognitive (Spatial) Tasks

Level 1	Understand the concept of "getting out"
Level 2	Understand the meaning of "in" and "out"
Level 3	Understand the concepts of "go in," "come out," and "under"
Level 4	Understand "inside," "outside," "underneath," and "on top of"
Level 5	Understand the meanings of "put it around" and "take it off"
Level 6	Besides what was learned before, understand one more meaning of "up"
Level 7	Besides what was learned before, understand one more meaning of "next to"
Level 8	Besides what was learned before, understand the meanings of "close to," "behind"

Figure B.6: The Timing of Cognitive Skill (Spatial) Tasks across Difficulty Levels



Note: Level 1: Underdtand the concept of "getting out", Level 8: Besides what learned before, understand the meanings of "close to", "behind."

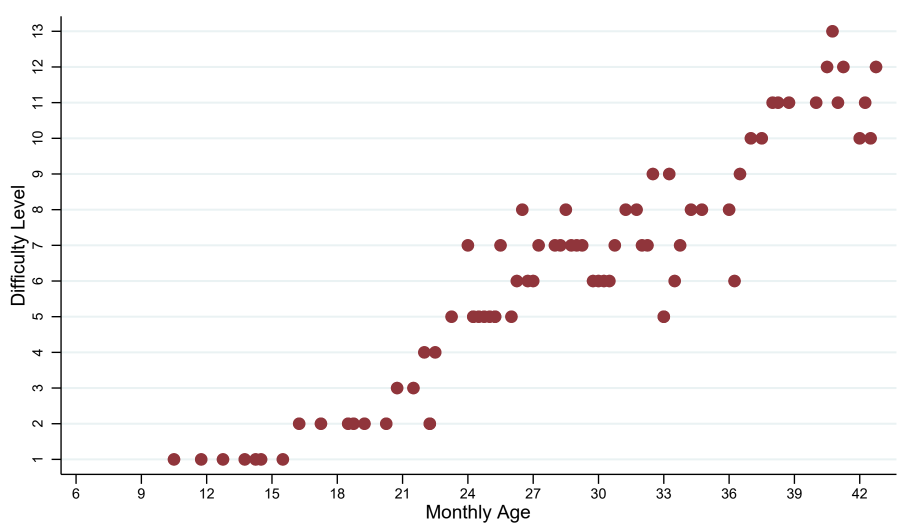
B.3.4.2 Knowing Objects and Objects' Functions

The knowing objects task set introduces preliteracy skills. It involves progressing interaction with pictures of objects and elements of storytelling. The tasks in the Cognitive Knowing Objects progress from a simple understanding of the concept of pictures by acknowledging with vocalizations, to using receptive (heard) language to identify certain pictures. Receptive language is a skill developed prior to an expressive language where a child forms words to communicate. The children must use their expressive language to complete the following tasks that increase with difficulty as they must develop more and more language to identify an increasing number of images. To progress through level 7 and beyond, the child must display an increasingly sophisticated understanding of the stories presented, first simply naming actions, then answering questions, then talking abstractly about the story. Levels 10, 11, 12, and 13 ask the child to take the information presented and build on it by discussing the uses of objects presented and making connections with other images.

Table B.8: Difficulty Level List for Cognitive (Understanding Objects) Tasks

Level 1	The child can look at the pictures and vocalize
Level 2	Name the objects and ask the child to point to the pictures accordingly
Level 3	The child can name the objects in one picture, and point to the named picture
Level 4	The child can name the objects in two or more pictures, and point to the named picture
Level 5	The child can point out named pictures, and say names of three or more
Level 6	The child can point out the picture mentioned, and correctly name the name of 6 or more pictures
Level 7	The child can talk about the pictures, answer questions, understand or names the verbs (eat, play, etc.)
Level 8	The child can follow the storyline, name actions and answer question
Level 9	The child can understand stories, and talk about the content in the pictures
Level 10	The child can keep up with the development of story
Level 11	The child can say the name of each graphic, discuss the role of each item, and then link the graphics in the card together
Level 12	The child can name the items in the picture, link the different pictures together, and discuss some of the activities in the pictures
Level 13	The child can name the things in the picture and talk about the function of objects

Figure B.7: The Timing of Cognitive Skill (Understanding Objects) Tasks across Difficulty Levels



Note: Level 1: Look at the pictures and vocalize; Level 13: The child can name the things in the picture and talk about the function of objects.

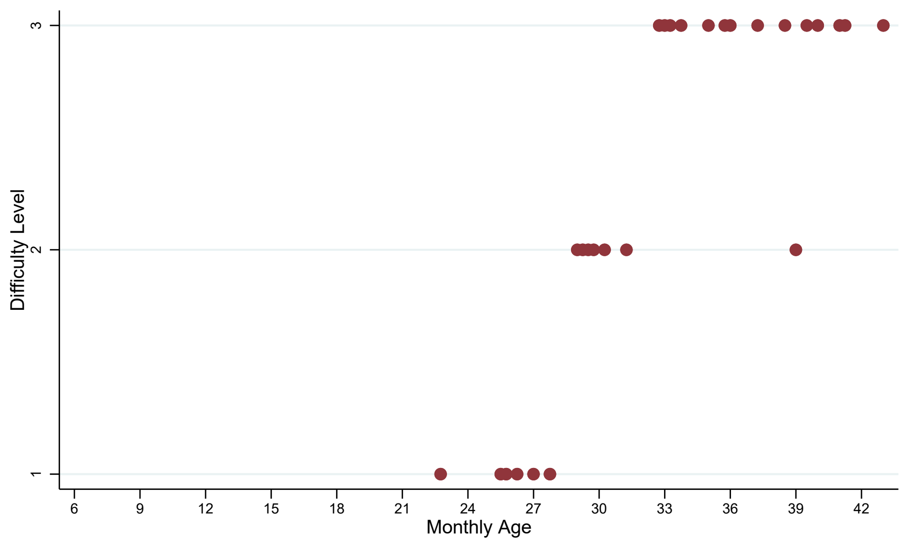
B.3.4.3 Color

In the color skill set, tasks progress from passive interactions (child hearing about color) to actively naming colors, to finally making connections with colors.

Table B.9: Difficulty Level List for Cognitive (Color) Tasks

Level 1	Caregiver talks about the color
Level 2	The child can identify the color
Level 3	Understand color and match different colors

Figure B.8: The Timing of Cognitive Skill (Understanding Color) Tasks across Difficulty Levels



Note: Level 1: Caregiver talks about color; Level 3: Understand color and match color with objects.

Table B.10: Difficulty Level List for Cognitive (Order: Understanding Upward, Forward, First, Some, All, Next, and Last) Tasks

Level 1	Child learns how to string beads and understands the meanings of "upward" and "downward"
Level 2	Understand the meanings of "upward," "downward," "first," and "then"
Level 3	Understand the concepts of "first," "finally," "in front of," and "behind"

Cognitive ability progresses into more abstract concepts of direction "upward" and "downward." Then, relative concepts of "first," "last," or "behind" are introduced.

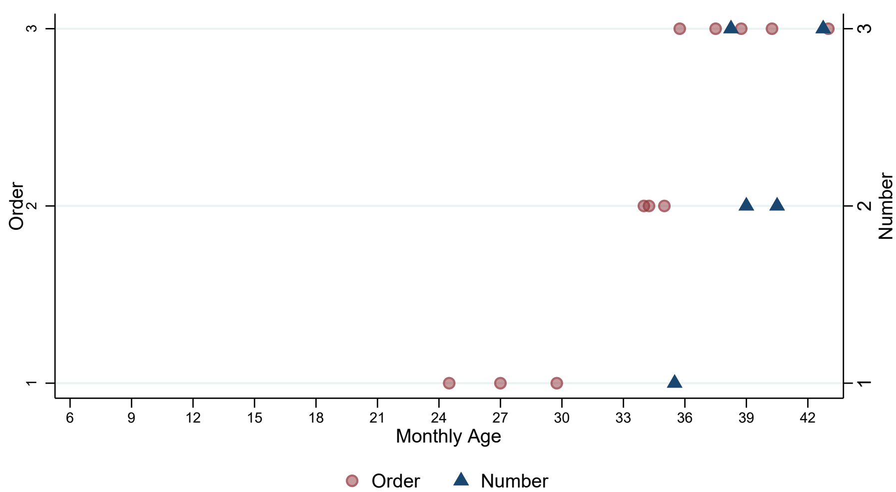
B.3.4.4 Number

Table B.11: Difficulty Level List for Cognitive (Number) Tasks

Level 1	Child learns how to count, can count up to 4
Level 2	Counting from 1 to 4, and then count two objects: 1, 2
Level 3	Children can count from 1 to 4 and sort the card by the number of points on each card

Number tasks progress from the learning of numbers in order to understanding one-to-one relationships of numbers to objects when counting. Finally, the concept of number representation is introduced.

Figure B.9: The Timing of Cognitive Skill (Understanding Order and Numbers) Tasks across Difficulty Levels



Note: Order level 1: Understanding "upward and downward"; level 3: Understanding the concepts of "first, finally, in front of, and behind."
Number level 1: learn how to count; level 3: Sorting the card by the number of points on each card.

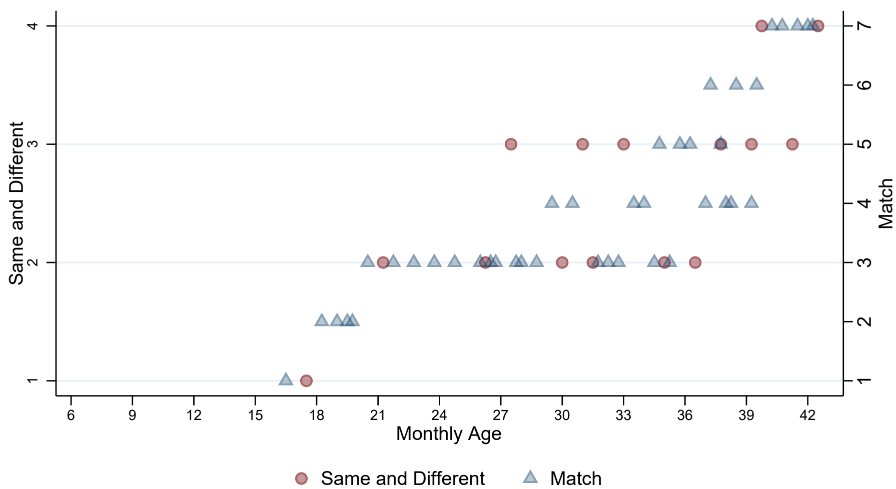
B.3.4.5 Match

These tasks consist of matching different pieces from simple puzzles to complicated puzzles. This set of tasks builds on the child's spatial awareness skills. The ability to fill in missing objects and understand how objects fit together is important in developing spatial awareness. The individual tasks progress from simply placing 1-2 puzzle pieces, completing the puzzle, making patterns, and using emerging language skills to describe pieces. As the children gain proficiency in these skills, they can complete puzzles of increasing complexity and restore the jumbled pieces to the original puzzle.

Table B.12: Difficulty Level List for Cognitive (Match) Tasks

Level 1	Put one piece into the puzzle
Level 2	The child is able to put at least two pieces in the puzzle
Level 3	The child can complete the simple puzzle
Level 4	The child can complete the puzzle and name different pieces
Level 5	The child learns to put together puzzle pieces to form the complete pattern
Level 6	With the caregiver's help, the child can complete the puzzle with more pieces
Level 7	The child can restore the puzzle to the original

Figure B.10: The Timing of Cognitive Skill (Matching and Understanding) Tasks across Difficulty Levels



Note: Same and Different level 1: teaching the meaning of "same", level 4 the child understands the concept of "same" and "different."
Match level 1: teaching one shape puzzle; level 7: match different shape pieces.

B.3.5 Language Skill

Language skill is the ability of children to communicate their needs, thoughts, feelings and ideas in a way that the caregiver can understand. It includes vocalizations, gestures, spoken words, and other signals.

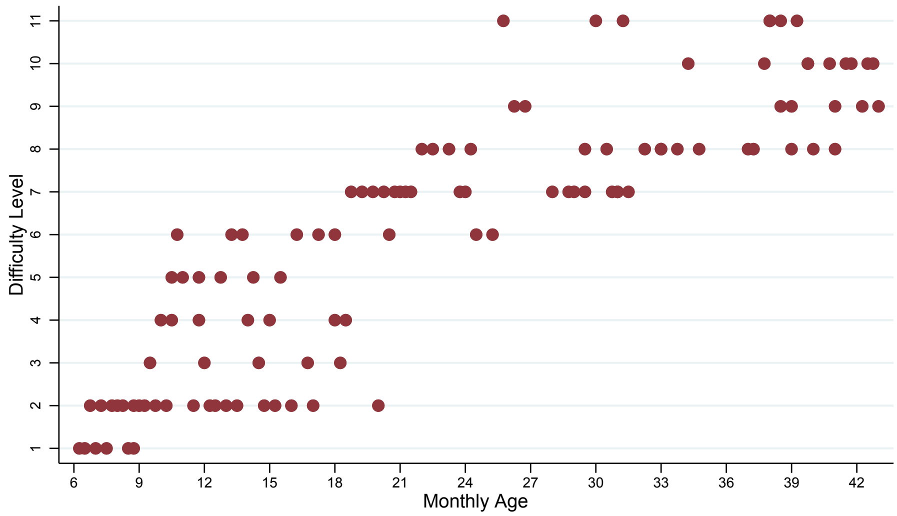
B.3.5.1 Learn words

Table B.13: Difficulty Level List for Language (Knowing Objects and Understanding Their Functions) Tasks

Level 1	Caregiver and baby make sounds to each other to interact
Level 2	Caregiver tells baby the things she does in the house
Level 3	To teach baby to recognize people's names
Level 4	Baby learns movements that show intimacy: clapping, bye-bye, and thank you
Level 5	Caregiver and child look at the pictures together, and let the child vocalize and touch the pictures
Level 6	Baby is to recognize at least one body part
Level 7	The child identifies and/or names ordinary objects
Level 8	The child points to the pictures which are being named, names one or more pictures, mimic the sound of the objects
Level 9	The child points to the pictures which are being named, names two or more pictures, mimic the sound of the objects
Level 10	The child points at 7 or more than 7 pictures and talk about them
Level 11	Teach the child some simple descriptive words and the child names objects at home, and tells the usage of those objects

The language skill tasks increase in difficulty with the expectation that the child will learn to identify and use expressive language to indicate understanding. The tasks begin with the baby passively listening as the caregiver makes sounds and speaks. The child then plays a more active role, expected to indicate understanding (receptive language) and use simple gestures to indicate meaning. As understanding and vocabulary increase, the child will name more pictures and learn to describe them. Finally, the child will learn the names and uses of objects in the child's everyday environment.

Figure B.11: The Timing of Language Skill (Knowing Objects) Tasks across Difficulty Levels



Note: level 1: Teach some new sounds; level 11: The child can name, discuss objects and mimic their sounds.

B.3.5.2 Dialogue

In this set of tasks, the caregiver talks to the children.

Table B.14: Difficulty Level List for Language (Dialogue) Tasks

Level 1	Caregiver talks to the baby when doing housework
Level 2	Use words that child learned to answer or create a new conversation

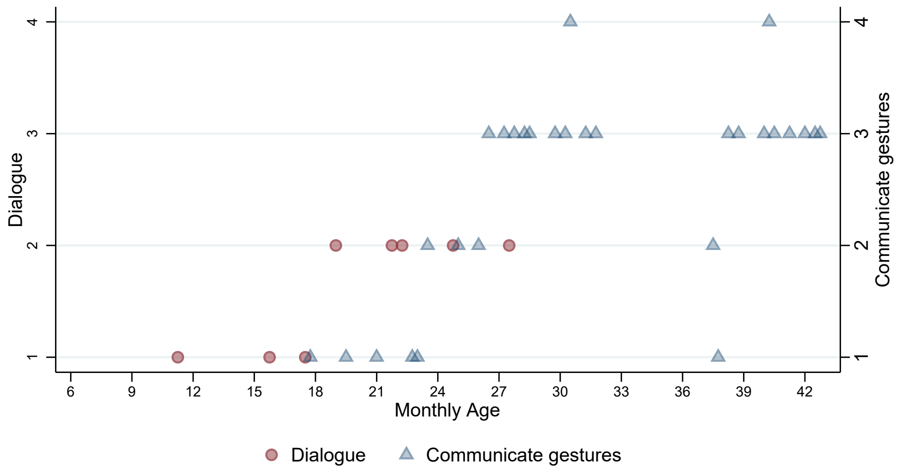
As the child grows, the caregiver progresses from simply narrating events to building on words the child has learned to scaffolding language development.

B.3.5.3 Communicate Gestures

Table B.15: Difficulty Level List for Language (Communicate Gestures) Tasks

Level 1	The baby listens to simple instructions given by the caregiver
Level 2	Caregiver performs some activities with the child
Level 3	Let the child learn to talk about the pictures, act according to the pictures, answer questions, and name related actions

Figure B.12: The Timing of Language Skill (Communicate Gestures) Tasks across Difficulty Levels



Note: Dialogue level 1: Talking to the child; level 2: Using the words the child learned to create conversation.
Communicate gestures level 1: the child listens to simple instructions; and level 4: the child can act as other roles, e.g., father, mother.

C Baseline Comparisons

C.1 Baseline Comparison for Children Enrolled in Jan 2015

In order to examine the quality of the randomization, in this section, we compare both the targeted and untargeted moments or distributions in the randomization design between the treatment and control groups. In Figure C.1, we give a comparison of the variables which are used in designing the matched pair. We can find that the treatment group and the control group have very similar distributions for the variables used for the randomization design. The Kolmogorov-Smirnov test p-values are all above 0.7, which indicates that we cannot reject the hypothesis that the distributions of treatment and control groups are identical.

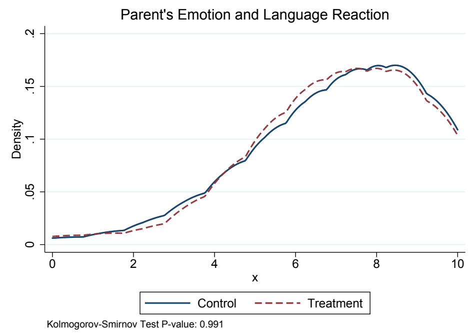
Next, we compare the variables which are not considered in the randomization estimation. We conduct this comparison for living conditions; family education levels, family structure, and economic conditions; pregnancy knowledge, pregnancy behavior, and the situations in pregnancy; children's health and development measures; and parent-child interaction.

Figure C.2 shows that well or spring water is the main water source for cooking in both treatment and control groups. About 95% of households have stable electricity for daily life. In Huachi county, there is a kind of traditional cave dwelling housing (Yaodong). 70% of households are still living in this kind of traditional housing. Figure C.3 shows the outward appearances of Yaodong in Huachi county. The fractions of different types of durable goods owned by each household are presented in Figure C.4. All t-test p-values are above 0.05, and there is no significant difference between treatment and control group households in terms of the ownership of durable goods. Almost every household has a cell phone and television. One notable fact is that about 70% of households have at least a motorcycle and above 20% of households own their cars. The ownership of an automobile is higher than 14%, which is the car parc rate in China. Most residents are living in a mountainous area. Cars or motorcycles are important tools to connect to places outside of the village.

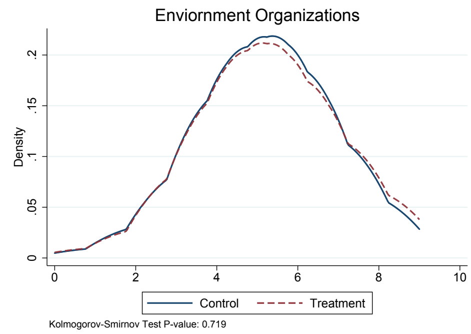
In Huachi county, the family structure is quite stable. For example, in Figure C.5 more than 98% of children's fathers and mothers are married or cohabitating. Figure C.6 shows the education distribution for different household members. The children's fathers and mothers have higher education levels than the grandfather-mother generation. More than 62% of children's fathers finished at least nine years of compulsory education. About 19% of the fathers graduate from high school or above. For the children's mothers, about 55% of them finished at least nine years of compulsory education, and about 11% are high school graduates or above. The fractions of mothers who have graduated primary school or middle school are similar to the fractions of the fathers. For the grandfather-mother generation, it is clear that grandmothers are less educated. More than 40% of them do not have any formal education.

Figure C.1: Distributions of Outcomes Used in Designing Matched Village Pairs

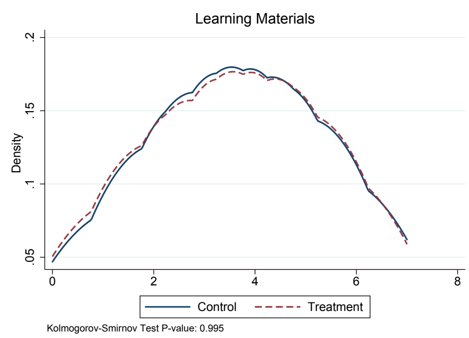
(a)



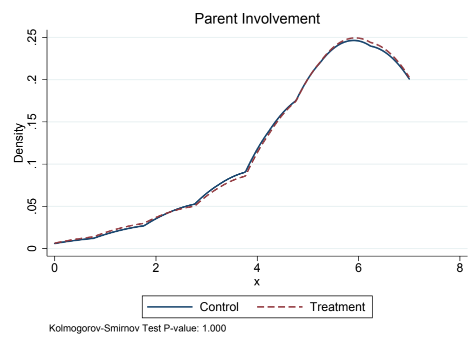
(b)



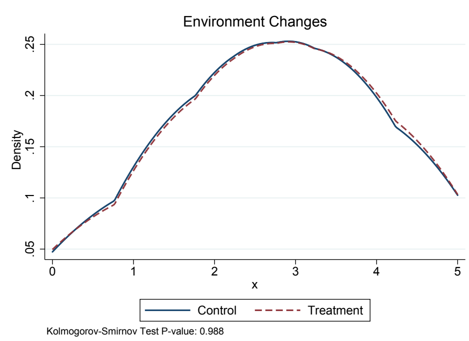
(c)



(d)



(e)



(f)

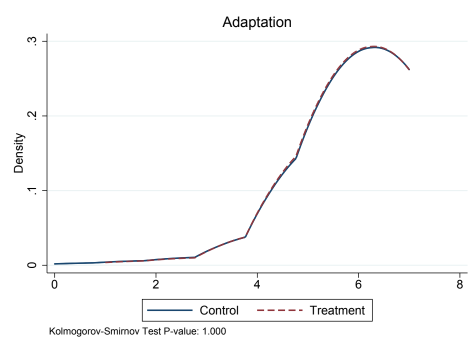
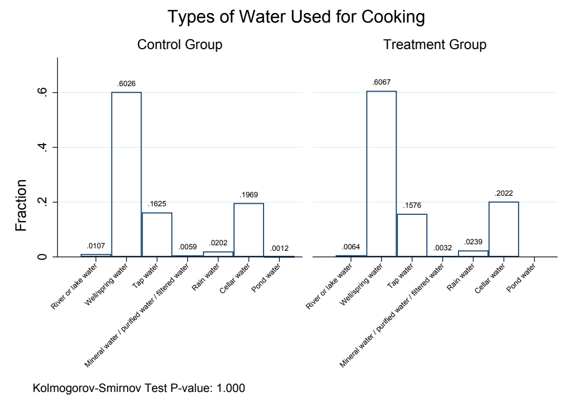
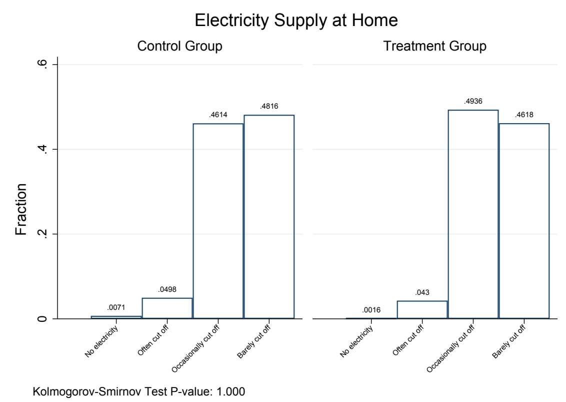


Figure C.2: Living Conditions

(a)



(b)



(c)

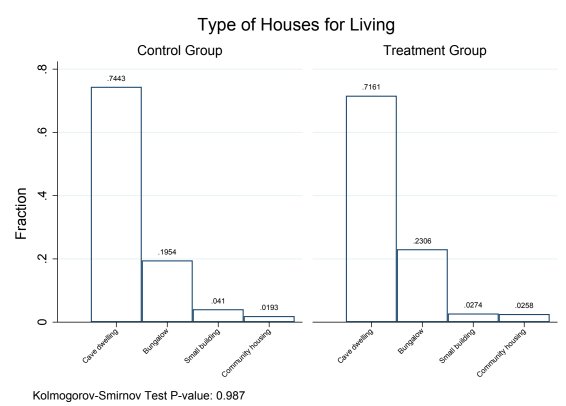


Figure C.3: Yaodong in Huachi County

(a)

(b)

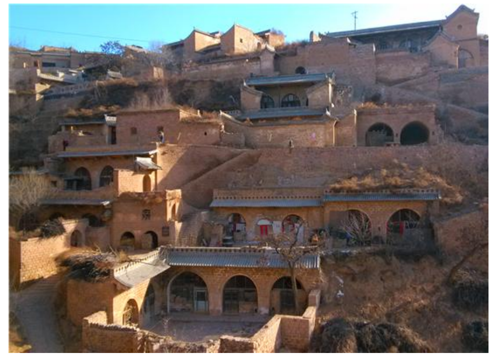
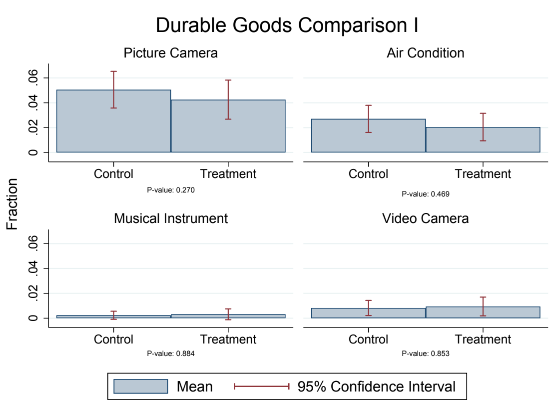
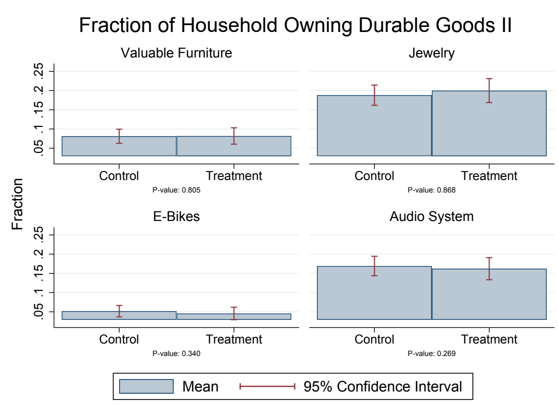


Figure C.4: Fraction of Households Owning Durable Goods

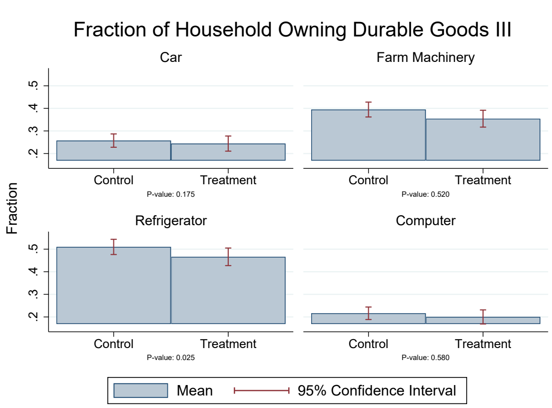
(a)



(b)



(c)



(d)

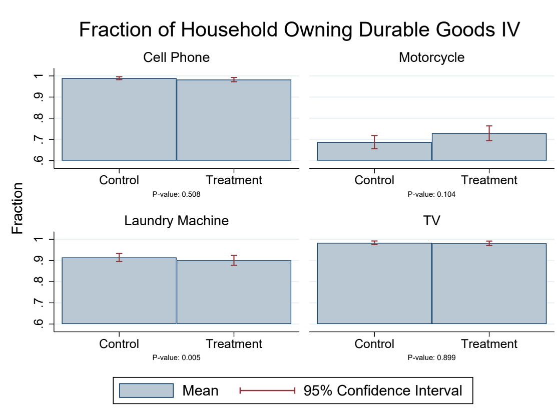
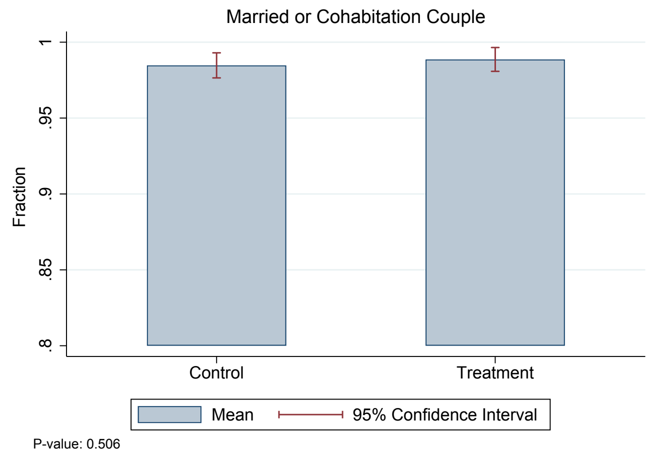


Figure C.5: Family Structure



In general, we can see that the education distributions for different household members between the control and treatment groups are very close. The Kolmogorov-Smirnov test shows that the education level distributions are identical between the control and treatment groups.

In Table C.1, we compare household annual income and consumption by categories. Since there are multiple income sources (e.g., wage income, agriculture income, and government subsidy) for the rural households, in the table we lay out main income sources of the control and treatment groups at the baseline. The column of "p-value" gives the statistics testing whether the mean values are different between the two groups. All p-values are greater than 5%, which means that we cannot reject the null hypothesis that the mean values of the two groups are equal to each other.

Figures C.7–C.9 give a summary of the knowledge of pregnancy and the performance in pregnancy. Figure C.7 provides the comparison of pregnancy knowledge between treatment and control groups. In general, control group individuals have greater knowledge of pregnancy but we cannot reject the null that both groups have equal knowledge of pregnancy. Figure C.8 shows the pregnancy behaviors which would affect children's health outcomes. We cannot find significant difference between treatment and control groups. Almost no mother smokes or drinks during her pregnancy. More than 80% of the mothers had prenatal check experiences, and 60% of the mothers had prenatal checks in the first three months of pregnancy. Also, we cannot find significant differences in the health conditions during pregnancy between treatment and control groups in Figure C.9.

Figure C.10 gives the duration of breastfeeding and the time of introducing complementary food to the infants. The distributions are very close for both treatment and control

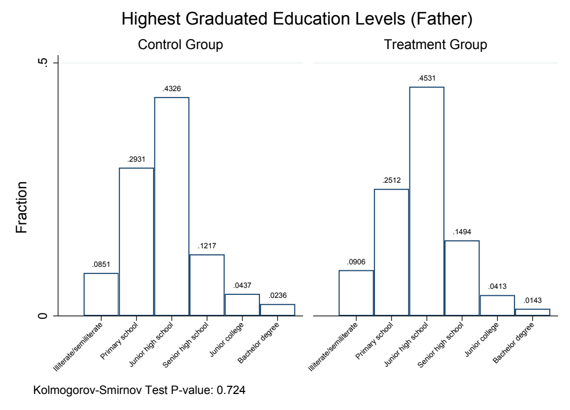
Table C.1: Consumption and Income Comparisons (Baseline)

Household Level	Control	Treatment	p-value
Agricultural Income	10284.66	7055.83	0.09
Standard Error	(1655.21)	(1912.39)	
Observations	704	515	
Government Subsidy	2780.42	2321.70	0.06
Standard Error	(220.42)	(244.40)	
Observations	751	567	
Remittance	15632.60	15969.94	0.87
Standard Error	(1233.56)	(2127.09)	
Observations	544	408	
Wage Income (After Tax)	34934.14	31255.06	0.39
Standard Error	(3102.52)	(4276.55)	
Observations	92	64	
Food Consumption	7861.47	9638.88	0.62
Standard Error	(1267.94)	(1956.49)	
Observations	703	513	
Total Consumption	42767.85	41796.35	0.92
Standard Error	(6504.43)	(10021.38)	
Observations	846	629	

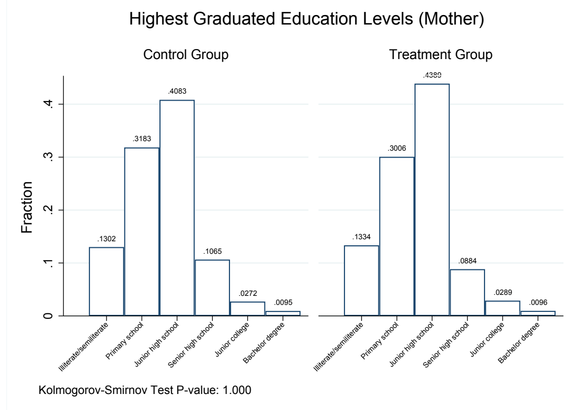
p-values are calculated by bootstrapping and clustering at the level of the randomized paired villages.

Figure C.6: Family Member Education Levels

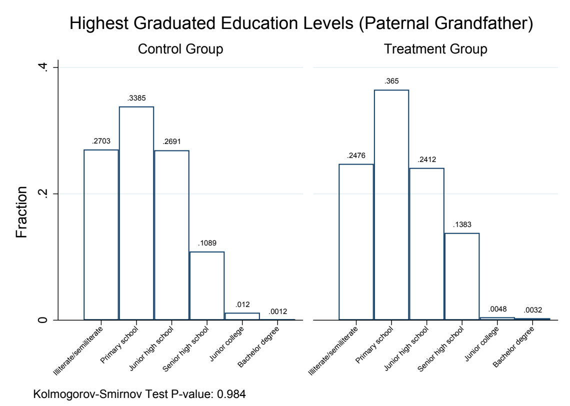
(a)



(b)



(c)



(d)

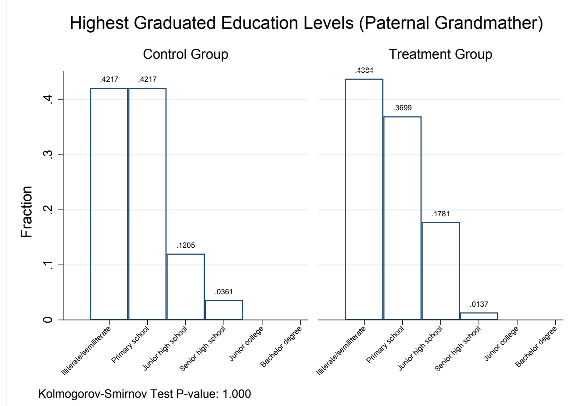
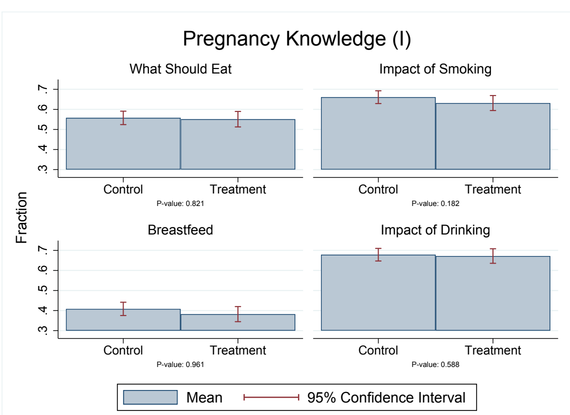


Figure C.7: Pregnancy Knowledge

(a)



(b)

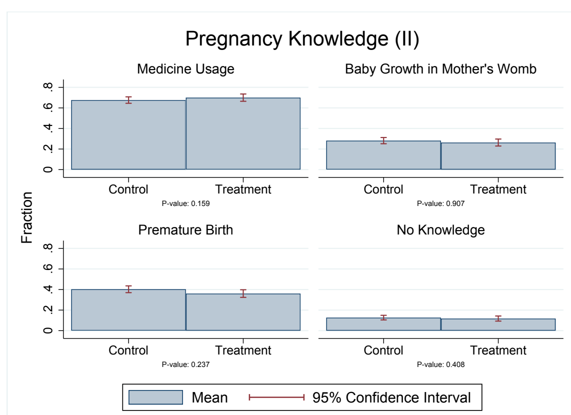
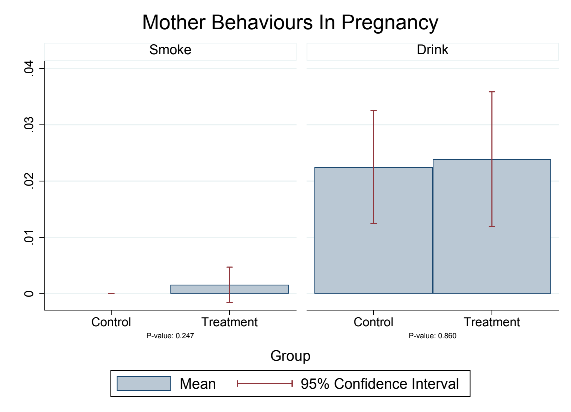


Figure C.8: Pregnancy Behavior

(a)



(b)

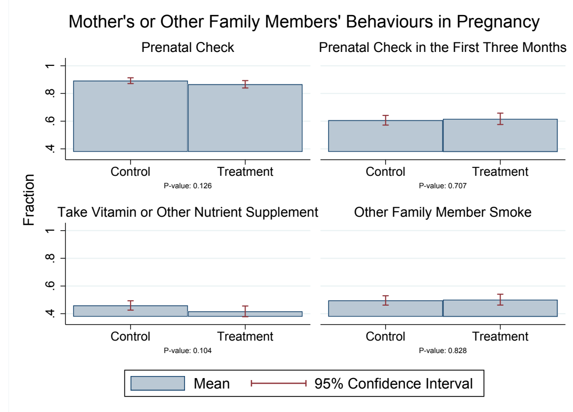


Figure C.9: Situations in Pregnancy

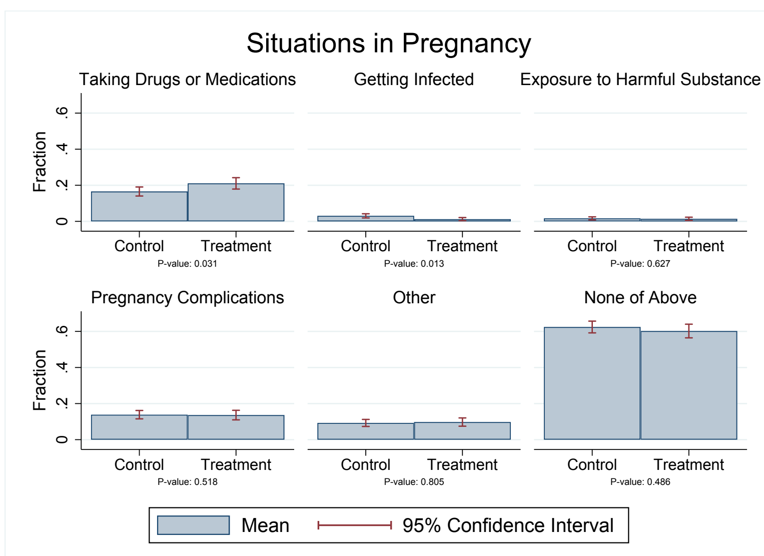
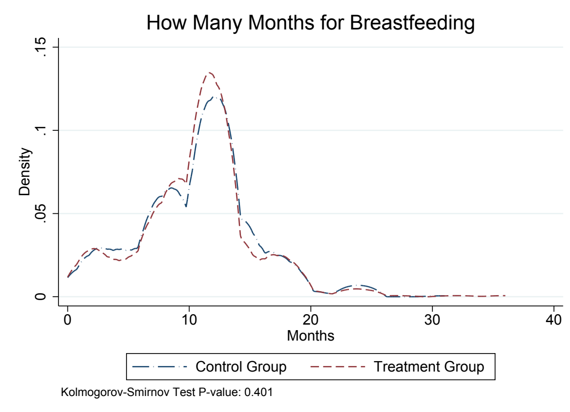


Figure C.10: Breastfeeding Behavior

(a)



(b)

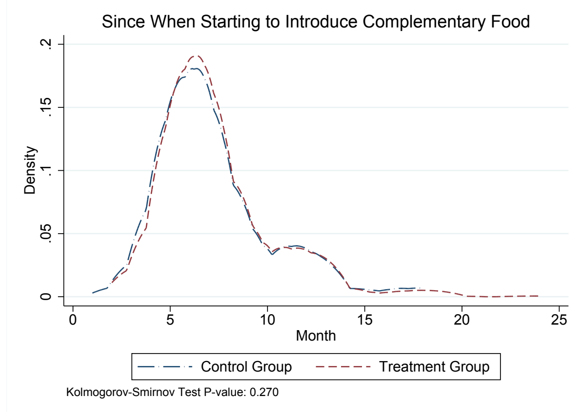
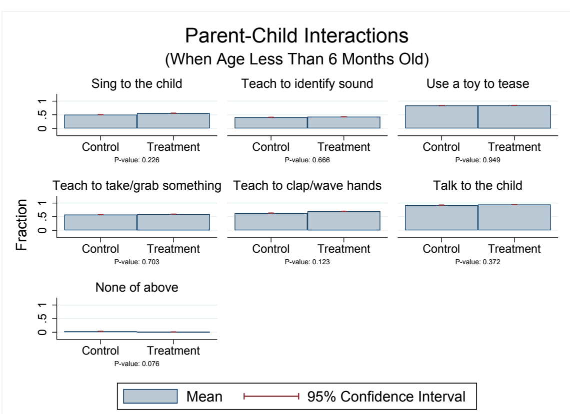
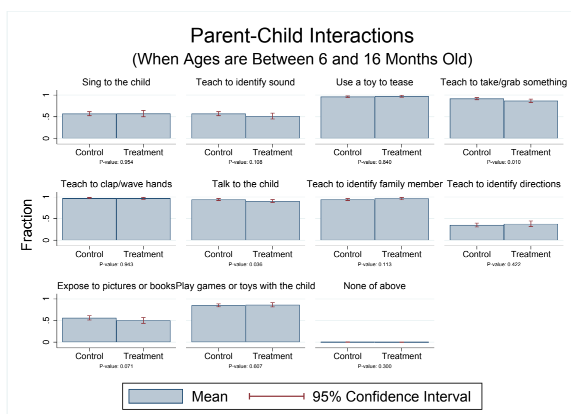


Figure C.11: Parent-Child Interaction

(a)



(b)



groups. From Figure C.11, we can find that the changes in parent-child interactions are related to the children's age. Also, there are no significant differences in both parent-child interactions between the two groups.

From the above comparisons, there are no significant differences in either target outcomes or the non-target variables between the control and treatment groups. In general, the randomization design works well in selecting matched pair villages.

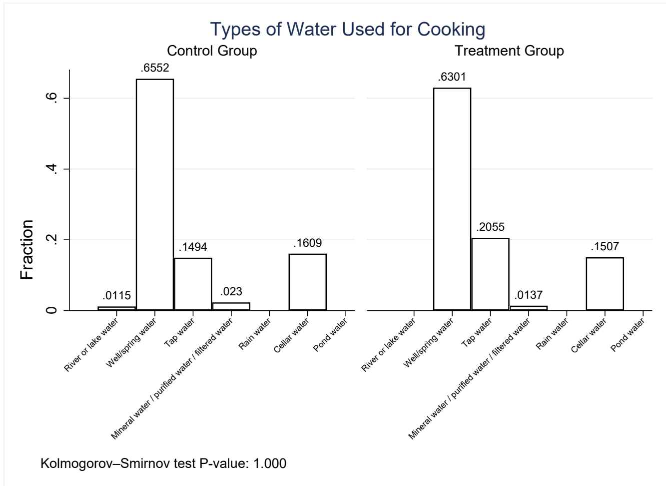
C.2 Baseline Comparison for Children Enrolled in Jan 2016

Similar to what was done in the above section, where we compared the baseline characteristics of both treatment and control groups, we now separately compare the same characteristics only with children who entered the sample during the second round of enrollment. The first round of enrollment started in January 2015 for children who were born between April 1, 2013 and November 30, 2014. However, eight months had passed by the time that the intervention formally began in September 2015, and some children from the initial enrollment aged out of the intervention age range. Meanwhile, more children were born and became eligible to participate, so 180 children were added to the study in January 2016. Therefore, the second round enrollment refers to this group of 172 children who were born between February 2015 and September 2015 (after attrition). Same as the first round of enrollment, the second round was again randomized into treatment and control group. More details on data and attrition can be found in Appendix G. We perform a baseline comparison to make sure that the treatment and control groups are balanced. In the following tables, we show that the characteristics have no statistically significant difference among this group of entrants.

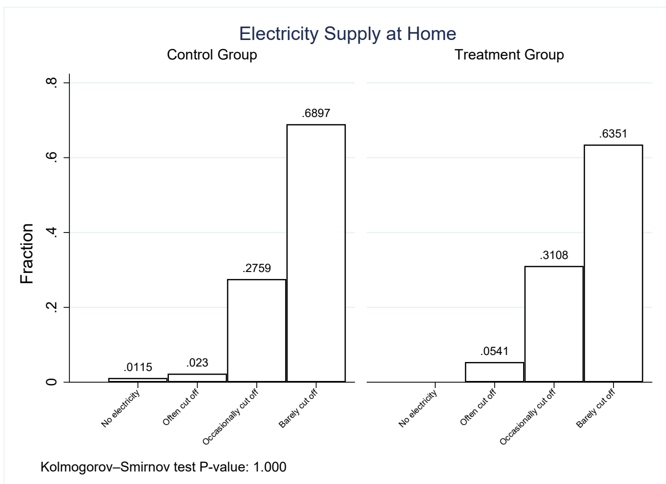
In order to examine the quality of the randomization, in this section, we compare both the targeted and untargeted moments or distributions between treatment and control groups. First, we examine the living conditions. Figure C.12 shows that well or spring water is the main water source for cooking in both treatment and control groups. More than 90% of households have stable electricity for daily life. About 70% of households live in the traditional cave dwelling housing, Yaodong. The Kolmogorov-Smirnov test p-values are all almost 1, which indicates that we cannot reject the hypothesis that the distributions of treatment and control groups are identical.

Figure C.12: Living Conditions

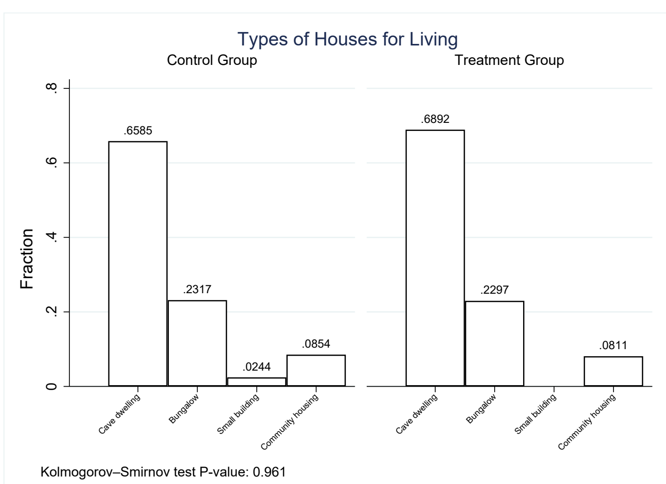
(a)



(b)



(c)



We can also see from Table C.2 that there is no significant difference between the treatment and control groups in regards to different types of income, government subsidy, or their expenditures.

Table C.2: Consumption and Income Comparisons (Baseline)

Household Level	Control	Treatment	*p-value*	Obs.
Agricultural Income	4564.225	4878.333	0.712	131
Government Subsidy	2497.610	2492.385	0.991	142
Remittance	20880.360	15566.670	0.239	112
Wage Income	37544.450	41250.000	0.774	21
Total Income	43653.010	38266.860	0.556	161
Total Consumption	38704.830	33817.570	0.431	161

The fractions of different types of durable goods owned by each household are presented in Table C.3. All t-test p-values are above 0.05, and there is no significant difference between treatment and control group households in terms of the ownership of durable goods. Almost every household has a cell phone and television. Almost no households own a video camera or any musical instrument. Over 60% of the households have a motorcycle.

Table C.3: Fraction of Households Owning Durable Goods (Baseline)

Household Level	Control	Treatment	*p-value*	Obs.
Car	0.264	0.329	0.354	170
E-Bikes	0.044	0.038	0.846	170
Motorcycle	0.648	0.633	0.835	170
Refrigerator	0.659	0.532	0.091	170
Laundry machine	0.879	0.861	0.724	170
TV	0.956	0.924	0.380	170
Computer	0.231	0.165	0.284	170
Audio system	0.242	0.190	0.417	170
Picture camera	0.033	0.038	0.861	170
AC	0.033	0.000	0.105	170
Cell phone	0.956	0.937	0.577	170
Valuable furniture	0.286	0.405	0.103	170
Musical instrument	0.000	0.013	0.284	170
Farm machinery	0.462	0.456	0.940	170
Jewelry	0.484	0.532	0.534	170

In Table C.4, we give a comparison of home environment. As we can see, the control group and the treatment group have very similar distributions of the variables.

Table C.4: Comparisons of Home Environment

	Control	Treatment	p-values	Observations
Environment Organizations	5.356	5.365	0.972	161
Learning Materials	4.253	3.73	0.052	161
Parent Involvement	5.839	5.689	0.366	161
Environment Changes	2.713	2.689	0.884	161
Adaptation	5.851	5.811	0.789	161

Parental involvement is based on the definition about the category of Infant-Toddler HOME Inventory: Parental Involvement including the following items: (1)Parent keeps child in visual range, looks at often. (2)Parent talks to child while doing household work. (3) Parent consciously encourages developmental advance. (4) Parent invites maturing toys with value via personal attention (5) Parent structures child's play periods, and (6) Parent provides toys that challenge child to develop new skills.

We also conduct this comparison for family structure, family education levels, living conditions, and economic conditions; pregnancy knowledge, pregnancy behavior, and the situations in pregnancy; children's health and development measures; and parent-child interaction. The family structure remains stable for the new entrants: 100% of children's fathers and mothers are married or in cohabitation. Figure C.13 shows the education distribution for different household members. The children's fathers and mothers generally have higher education levels than the grandfather-mother generation. More than 65% of children's fathers finished at least nine years of compulsory education. About 17% of the fathers graduate from high school or above. For children's mothers, about 58% of them finished at least nine years of mandatory education, and about 18% are high school graduates or above. The fractions of mothers who have graduated primary school or middle school are similar to the fractions of the fathers. For the grandfather-mother generation, it is clear that grandmothers are less educated. More than 40% of them do not have any formal education. In general, we can see that the education distributions for different household members between treatment and control groups are very close. The Kolmogorov-Smirnov test shows that the distributions of education levels are identical between control and treatment groups.

Figure C.13: Family Member Education Levels

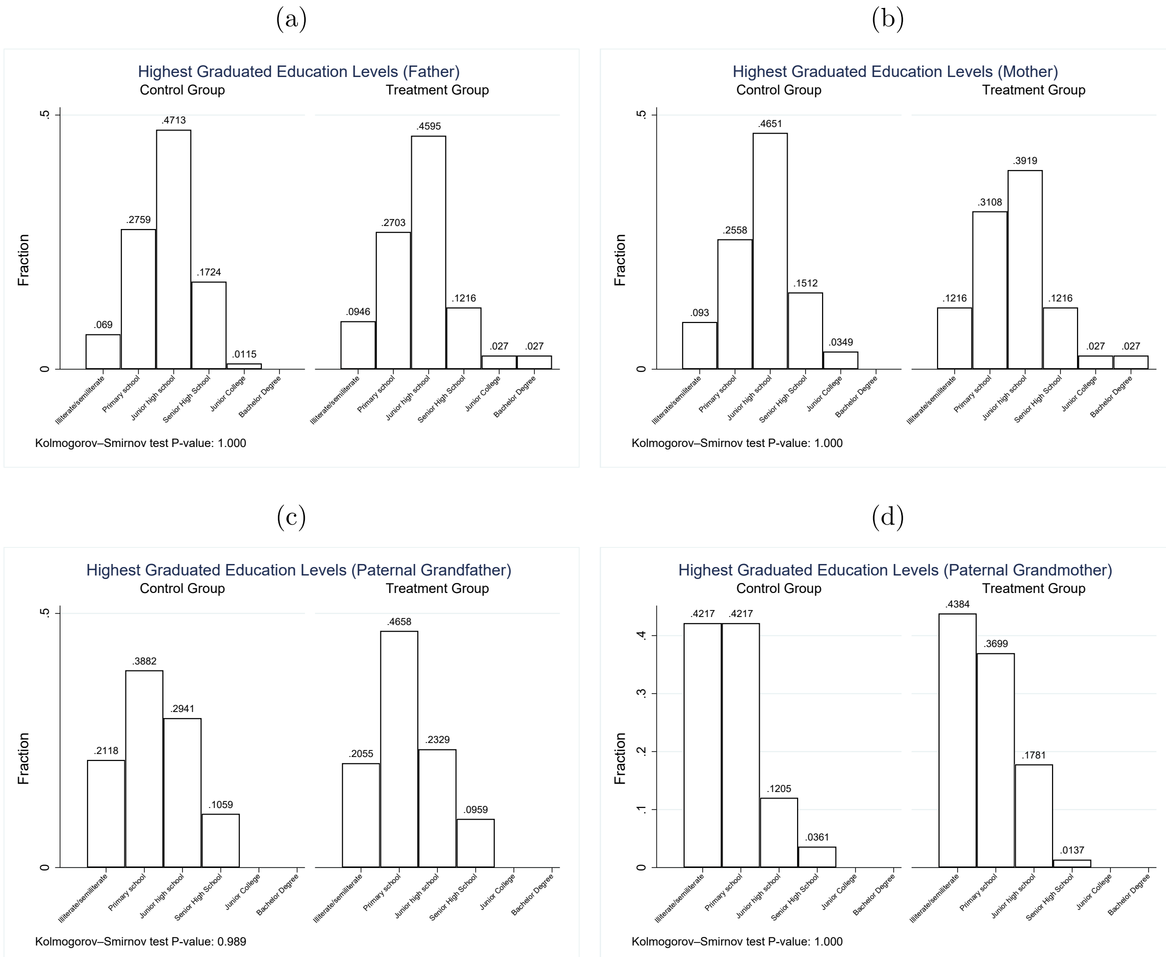


Table C.5 gives a summary of the knowledge of pregnancy. We cannot reject the null that both groups have equal knowledge of pregnancy. Table C.6 shows the pregnancy behaviors which would affect child health outcomes. We cannot find significant difference between treatment and control groups. Almost no mother smokes or drinks during her pregnancy. More than 90% of the mothers had prenatal check experiences, and over 60% of the mothers had prenatal checks in the first three months of pregnancy. Again, no statistical significance was found. Also, we cannot find significant differences in the health conditions during pregnancy between the control and treatment groups in Table C.7.

Table C.5: Pregnancy Knowledge (Baseline)

Whether they know	Control	Treatment	p-value	Obs.
What should eat	0.560	0.544	0.834	170
Impact of smoking	0.626	0.608	0.803	170
Breastfeed	0.451	0.367	0.273	170
Impact of drinking	0.571	0.671	0.185	170
Medicine usage	0.648	0.696	0.511	170
Baby growth in mother's womb	0.231	0.291	0.373	170
Premature birth	0.242	0.215	0.683	170
No knowledge	0.143	0.127	0.759	170

Table C.6: Pregnancy Behavior (Baseline)

Whether they	Control	Treatment	p-value	Obs.
Prenatal Check	0.953	0.946	0.842	159
Prenatal Check in the First 3 Mo	0.793	0.652	0.054	151
Take Vitamin/Supplement	0.463	0.432	0.700	156
Smoking	0.000	0.014	0.288	158
Drinking	0.000	0.014	0.285	159

Table C.7: Pregnancy Situation (Baseline)

Whether they	Control	Treatment	p-value	Obs.
Got infected	0.011	0.000	0.353	170
Exposure to harmful substance	0.011	0.025	0.482	170
Pregnancy complications	0.319	0.253	0.350	170
Other	0.044	0.076	0.380	170
None of above	0.462	0.468	0.930	170

From Table C.8, we find that there are no significant differences in parent-child interactions between the two groups. Table C.9 shows the comparison of child development measures between treatment and control groups at the baseline, and we find no significant difference between the two groups for different types of skills. Table C.9 also compares the birth weight and birth height, which again are not significantly different.

Table C.8: Parent-Child Interaction (Baseline)

Parent-Child Interaction	Control	Treatment	p-value	Obs.
Sing to the child	0.560	0.418	0.064	170
Teach to identify sound	0.538	0.557	0.810	170
Use a toy to tease	0.912	0.924	0.779	170
Teach to take/grab	0.890	0.873	0.738	170
Teach to clap/wave	0.868	0.886	0.725	170
Talk to child	0.868	0.911	0.375	170
Teach to identify family member	0.802	0.861	0.314	170

Table C.9: Children's Development (Baseline)

Children's Development	Control	Treatment	p-value	Obs.
Personal/Social	7.214	7.135	0.656	144
Fine Motor Adaptive	12.214	12.297	0.588	144
Language	11.157	11.027	0.491	144
Gross Motor	11.514	11.135	0.048	144
Denver Total	42.1	41.595	0.309	144
Birth Height	49.973	50.305	0.354	148
Birth Weight	49.973	50.305	0.354	148
Monthly Age	5.977	5.509	0.08	169

D Denver II Test

In Figures D.1 and D.2, we present the Denver test implemented during the intervention for both English and Chinese versions. The items are over 99% consistent between the two versions.

Figure D.1: English Denver Test

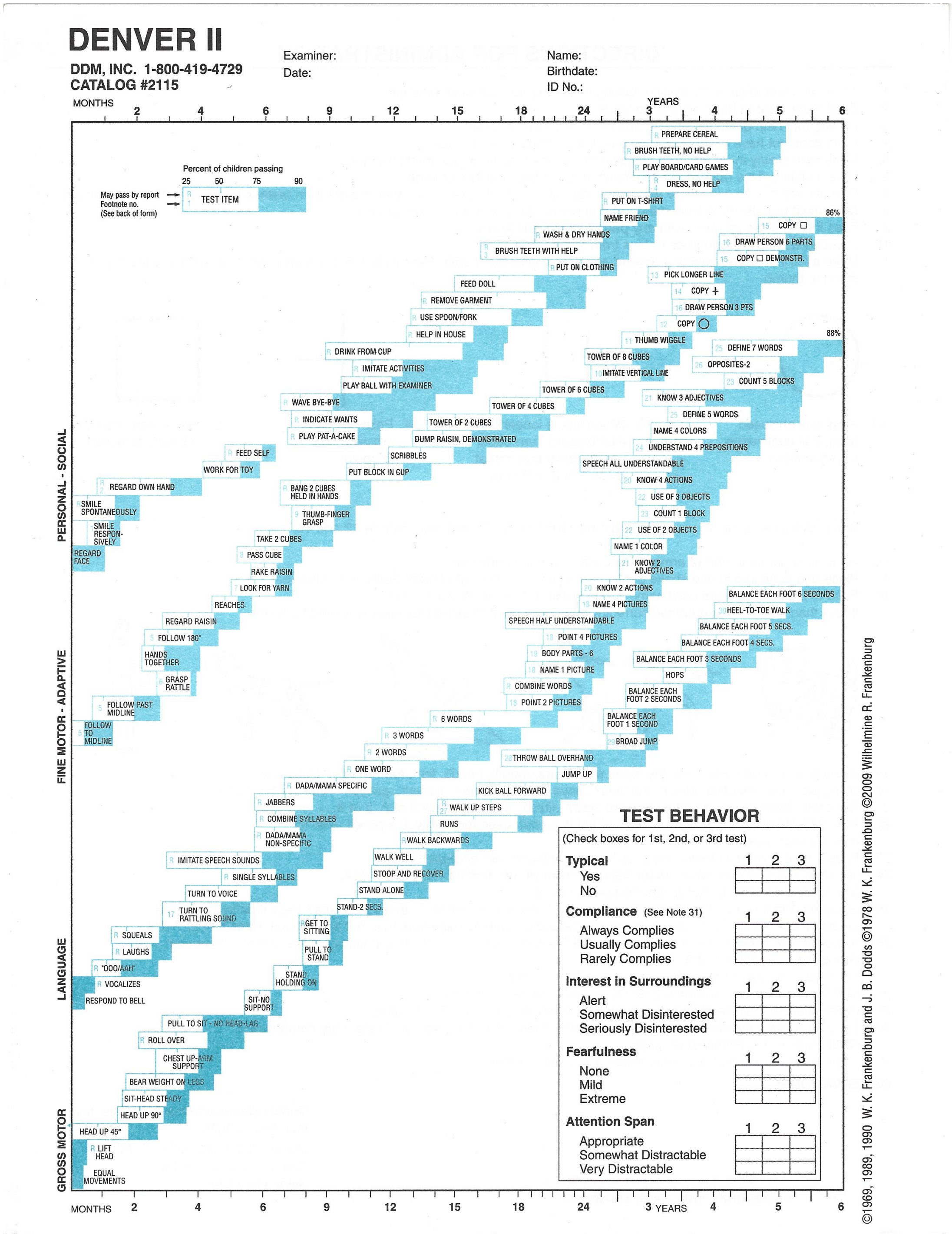
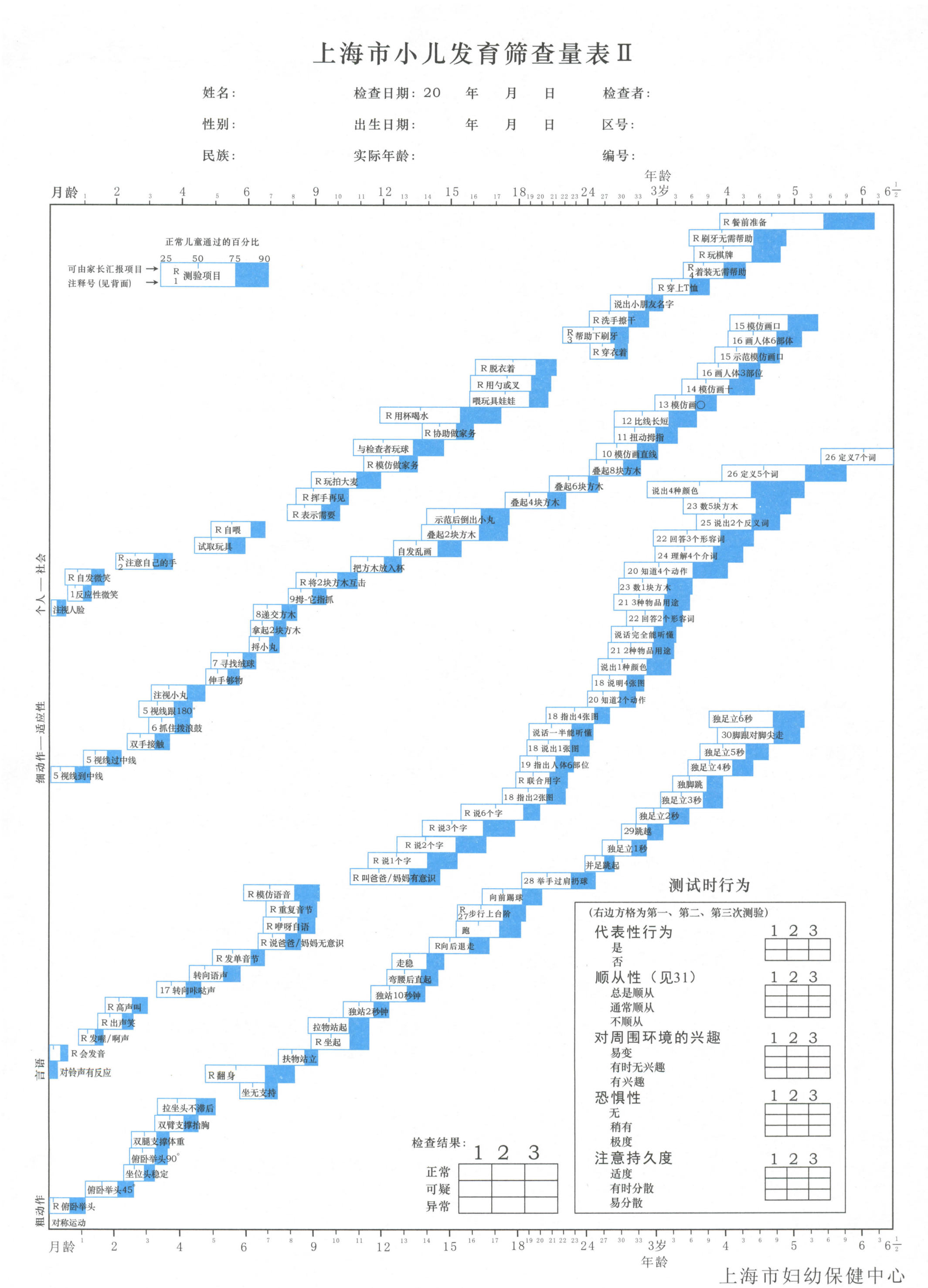


Figure D.2: Chinese Denver Test



E Linear Model Estimates on Raw Scores

Table E.1: Treatment Effects on Raw Scores

	(1) All	(2) All	(3) Children ≤ 2 Yrs at Enrollment	(4) All	(5) Children ≤ 2 Yrs at Enrollment
			Midline		
Language and Cognitive	0.533***	0.569***	0.689***	0.634***	0.754***
	[0.162, 0.895]	[0.161, 0.969]	[0.299, 1.090]	[0.234, 1.036]	[0.347, 1.173]
Fine Motor	0.064	0.166	0.200	0.195	0.228
	[-0.104, 0.233]	[-0.075, 0.412]	[-0.033, 0.444]	[-0.052, 0.467]	[-0.014, .488]
Social-Emotional	0.206**	0.274***	0.259***	0.285***	0.271***
	[0.044, 0.372]	[0.094, 0.452]	[0.073, 0.453]	[0.115, 0.463]	[0.067, 0.477]
Gross Motor	-0.140	-0.121	-0.009	-0.119	-0.031
	[-0.391, 0.110]	[-0.398, 0.156]	[-0.277, 0.276]	[-0.391, 0.148]	[-0.295, 0.243]
			Endline		
Language and Cognitive	1.031***	0.966***	1.172***	1.041***	1.247***
	[0.599, 1.472]	[0.509, 1.427]	[0.735, 1.591]	[0.601, 1.489]	[0.813, 1.687]
Fine Motor	0.224	0.205	0.232**	0.238*	0.265**
	[-0.006, 0.457]	[-0.021, 0.424]	[0.023, 0.454]	[0.009, 0.472]	[0.019, 0.530]
Social-Emotional	-0.133	-0.159	-0.093	-0.136	-0.066
	[-0.299, 0.038]	[-0.342, 0.022]	[-0.260, 0.071]	[-0.319, 0.051]	[-0.240, 0.109]
Gross Motor	0.085	0.106	0.101	0.112	0.122
	[-0.244, 0.422]	[-0.184, 0.405]	[-0.190, 0.396]	[-0.174, 0.401]	[-0.166, 0.405]
Pre-treatment Covariates	No	No	No	Yes	Yes
IPW	No	Yes	Yes	Yes	Yes

1. The 95% confidence intervals in parentheses are constructed by wild bootstrap clustered at the village level.

2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table E.2: Treatment Effects on Raw Scores

(Female)

	(1) All	(2) All	(3) Children ≤ 2 Yrs at Enrollment	(4) All	(5) Children ≤ 2 Yrs at Enrollment
			Midline		
Language and Cognitive	0.368	0.365	0.479**	0.388	0.518**
	[-0.089, 0.813]	[-0.105, 0.842]	[0.020, 0.939]	[-0.073, 0.882]	[0.100, 0.996]
Fine Motor	0.135	0.132	0.183	0.156	0.192
	[-0.110, 0.379]	[-0.113, 0.369]	[-0.0337, 0.396]	[-0.103, 0.418]	[-0.030, 0.411]
Social-Emotional	0.352**	0.348**	0.368***	0.382***	0.399***
	[0.098, 0.617]	[0.101, 0.584]	[0.125, 0.609]	[0.126, 0.640]	[0.132, 0.663]
Gross Motor	-0.043	-0.078	-0.034	-0.069	-0.055
	[-0.363, 0.284]	[-0.421, 0.276]	[-0.372, 0.298]	[-0.427, 0.303]	[-0.409, 0.318]
			Endline		
Language and Cognitive	0.775*	0.827*	0.859*	0.869*	0.894*
	[-0.118, 1.611]	[0.022, 1.618]	[0.007, 1.683]	[0.004, 1.728]	[0.073, 1.749]
Fine Motor	0.350	0.313	0.348	0.347	0.369
	[-0.044, 0.790]	[-0.043, 0.705]	[-0.004, 0.698]	[-0.066, 0.797]	[-0.040, 0.807]
Social-Emotional	-0.147	-0.169	-0.146	-0.164	-0.148
	[-0.339, 0.045]	[-0.363, 0.033]	[-0.363, 0.066]	[-0.333, 0.011]	[-0.325, 0.036]
Gross Motor	0.167	0.209	0.236	0.208	0.223
	[-0.434, 0.724]	[-0.346, 0.779]	[-0.347, 0.804]	[-0.321, 0.728]	[-0.319, 0.739]

1. The 95% confidence intervals in parentheses are constructed by wild bootstrap clustered at the village level.

2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table E.3: Treatment Effects on Raw Scores

(Male)

	(1) All	(2) All	(3) Children $\leq$ 2 Yrs at Enrollment	(4) All	(5) Children $\leq$ 2 Yrs at Enrollment
			Midline		
Language and Cognitive	0.716***	0.807***	0.905***	0.853***	0.917***
	[0.204, 1.250]	[0.245, 1.378]	[0.354, 1.467]	[.306, 1.388]	[0.297, 1.520]
Fine Motor	0.046	0.193	0.208	0.211	0.231
	[-0.198, 0.304]	[-0.135, 0.551]	[-0.101, 0.539]	[-0.117, 0.590]	[-0.096, 0.604]
Social-Emotional	0.162	0.240	0.212	0.229	0.188
	[-0.083, 0.411]	[-0.012, 0.499]	[-0.071, 0.485]	[-0.044, 0.500]	[-0.101, 0.480]
Gross Motor	-0.154	-0.086	0.041	-0.126	-0.020
	[-0.437, 0.123]	[-0.447, 0.271]	[-0.277, 0.368]	[-0.426, 0.193]	[-0.306, 0.289]
			Endline		
Language and Cognitive	1.198***	0.948**	1.273***	1.037***	1.376***
	[0.548, 1.822]	[0.233, 1.635]	[0.724, 1.837]	[0.375, 1.730]	[0.766, 1.980]
Fine Motor	0.138	0.111	0.128	0.108	0.124
	[-0.118, 0.395]	[-0.151, 0.391]	[-0.132, 0.386]	[-0.146, 0.368]	[-0.136, 0.398]
Social-Emotional	-0.146	-0.194	-0.095	-0.181	-0.089
	[-0.391, 0.115]	[-0.456, 0.065]	[-0.315, 0.144]	[-0.479, 0.121]	[-0.369, 0.216]
Gross Motor	-0.060	-0.067	-0.066	-0.077	-0.059
	[-0.289, 0.171]	[-0.295, 0.168]	[-0.297, 0.156]	[-0.283, 0.134]	[-0.269, 0.161]

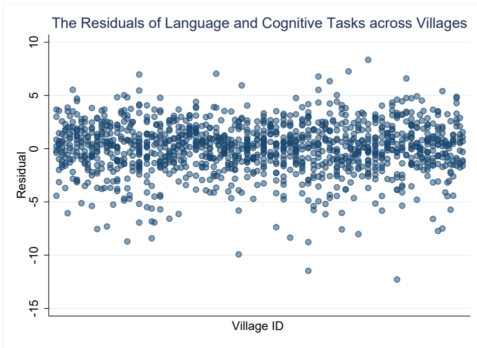
1. The 95% confidence intervals are constructed by wild bootstrap clustered at the village level.

2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

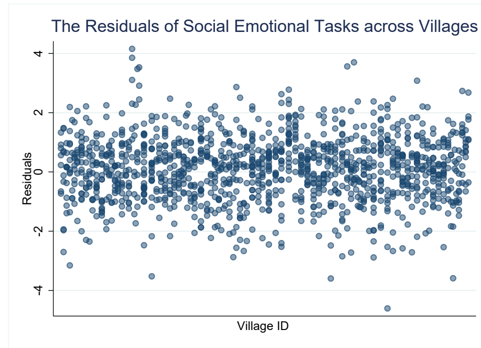
Figure E.1 shows that the residuals from these regressions are at best weakly correlated across villages.

Figure E.1: Test of Residual Independence across Villages

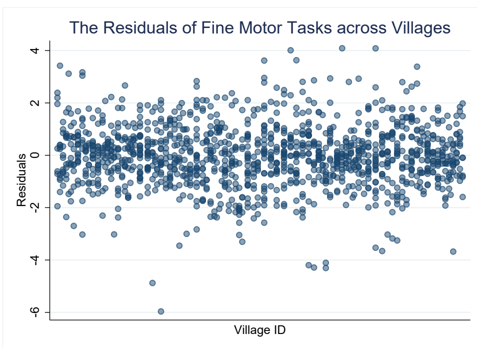
(a)



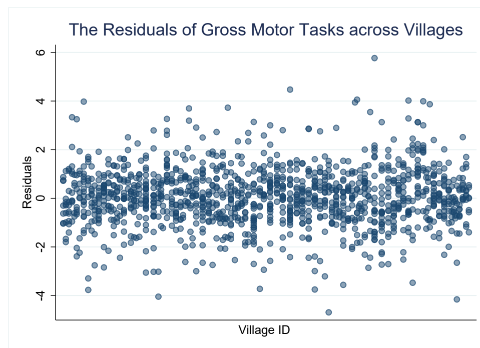
(b)



(c)



(d)



F Wild Bootstrap Procedure

The Cameron et al. (2008) procedure:

1. From OLS estimation on the original sample, we obtain the estimates $\hat{\boldsymbol{\beta}}$ and the CRVE $\hat{\Omega}$. Also, based on the null hypothesis $\boldsymbol{a}'\boldsymbol{\beta} = \mathbf{0}$, we reestimate the model to obtain restricted estimates $\tilde{\boldsymbol{\beta}}$ and residuals $\tilde{\boldsymbol{u}}$, calculate the cluster robust t statistic t_o

2. Do B iterations of this step. On the bth iteration:

   (a) Form a sample of V clusters $(\hat{\boldsymbol{y}}^*_{\boldsymbol{1}}, \boldsymbol{X_1}, \cdots, \hat{\boldsymbol{y}}^*_{\boldsymbol{V}}, \boldsymbol{X_V})$ by the following method. For each cluster v, $\boldsymbol{u_v^{*b}} = w_v^{*b}\tilde{\boldsymbol{u}}_{\boldsymbol{v}}$ and the w_v^{*b} are independent realizations of an auxiliary random variable w^* with zero mean and unit variance; then form $\boldsymbol{y^{*b}} = \boldsymbol{X'}\tilde{\boldsymbol{\beta}} + \boldsymbol{u_v^{*b}}$

   (b) Calculate the bootstrap estimates $\hat{\boldsymbol{\beta}}^{*^{\boldsymbol{b}}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y^{*b}}$ and the bootstrap covariance matrix and the bootstrap t-statistic t^{*b}

3. Reject H_0 at the level α if and only if $t_o < t^*_{[\alpha/2]}$ or $t_o > t^*_{[1-\alpha/2]}$

G Data, Attrition, and Nonresponse

This section documents data collection procedures, data attrition problems, and how we address data attrition problems. In January 2015, CDRF collected baseline information in Huachi county; 1,566 children were presented at that time. The RCT design was conducted based on the 1,566 children's survey information and village level administrative data, in which 796 children are in the treated villages and 770 children are in the control villages.

In September 2015, the home visiting intervention started. There was an eight-month gap between the baseline data collection and the first home visit. The local field team made two modifications to the original protocol before they started the first home visit. The first modification is that they included 76 new children in the intervention who were not surveyed in January 2015 but were eligible in September 2015 and, in addition, they excluded most of the children who were older than two years old in September 2015 (about 150 children).[7] The second modification is that they excluded children with urban hukou (around 90). Therefore, after the two adjustments, in September 2015, the sample size was 1,395, including 634 children in the treatment group and 761 children in the control group. In January 2016, 180 children from the younger cohort were added: 89 in the treatment group and 91 in the control group.[8] The data we use in our analysis include 1,567 observations: 1,395 of which were tracked since January 2015, plus 172 children for whom the baseline was January 2016 (8 children are missing from 180 samples). Finally, the sample we use includes 1,567 children, of which 715 are in the treatment group and 852 are in the control group.[9]

Table G.1 summarizes the sample created before the first intervention. The two main modifications targeted only the treatment group, hence most modifications came from the treatment group children. Since there were newly enrolled and also excluded children during this process, we examine the baseline comparison (the final sample with 1,567 children) between treatment and control groups in Section C, and find that there is no significant difference between treatment and control groups. When the field group started the home visits, there were 715 children on their name list. Among these 715 children, 705 children participated in the home visits (i.e., compliance rate is above 98%).

To have robust estimates, in the estimation, we also account for the sample adjustment in September 2015. According to the timeline, there were three stages at which data attrition

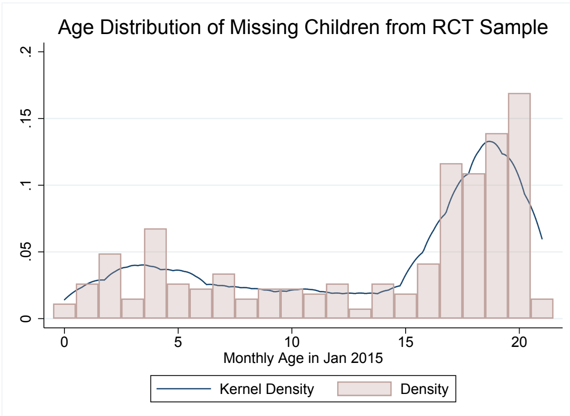
[7] For these 76 children, the field team collected their baseline information at the midline annual evaluation based on the parents retrospective responses.

[8] For these new children, they are a younger cohort from Huachi county and also take the Denver test assessment.

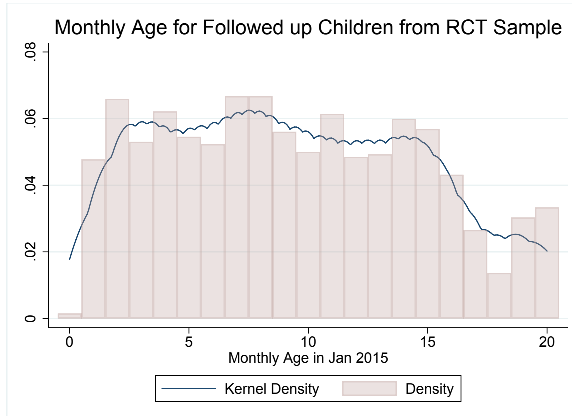
[9] Here, the reason why the final sample is 1,567 and not 1,566 is that we find two children share the same ID in the Denver test. We also find both children's information in the weekly home visit records. Therefore, we include the additional child.

Figure G.1: Age Distribution for the Samples before February 2016

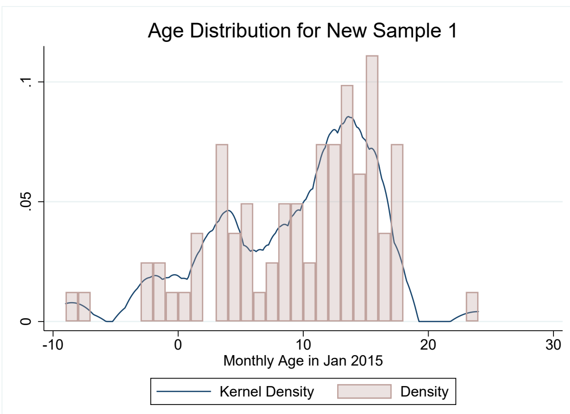
(a)



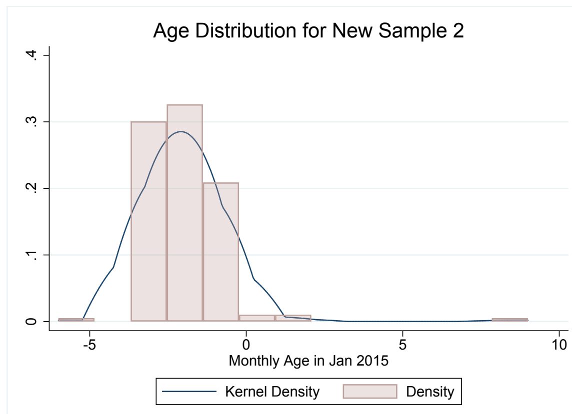
(b)



(c)



(d)



(e)

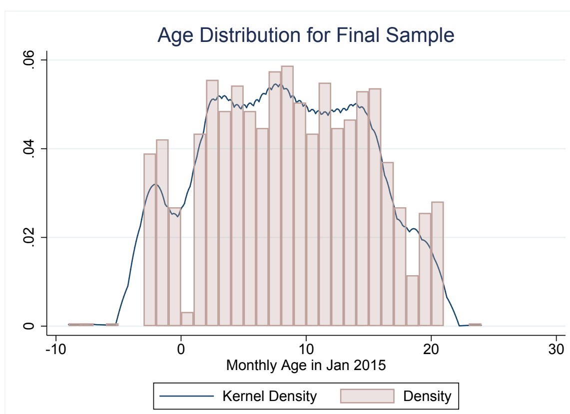


Table G.1: Huachi County Data Sample Before the Intervention

(January 2015-January 2016)

	Total	Treatment	Control
Baseline (January 2015)	1566	796	770
Adjustment in September 2015			
Adjustment 1 (include the children not surveyed in January 2015)	76	76	0
Adjustment 2 (exclude > 2 years old and urban hukou children)	-247	-238	-9
New Enrollment in January 2016	180	89	91
Missing	-8	-8	0
Final Sample	1567	715	852

occurred. The first stage is data attrition before February 2016. Before February 2016, 1,822 children were enrolled in the program, which included 1,566 from January 2015, 76 from September 2015, and 180 from January 2016. After January 2016, 245 children were not followed up with. For the home visiting intervention, the final list was based on the 715 children in the treatment group previously discussed, of whom 705 had at least one home visit (close to 99% of children in the treatment group had been treated). By July 2017, the average number of home visits was 74.

Since both the midline and endline child development assessments were conducted in a short time window (e.g., two weeks), and the annual review assessments were conducted in the town center hospital, data attrition appeared at both rounds of assessments. Table G.1 shows the decomposition of the 1,567 children in the followed-up sample. At the midline, there were 1,301 children who attended the Denver test (i.e., 636 in the treatment group, and 633 in the control group); for family survey information, information on more than 1,430 children was recorded. At the endline, there were 1,073 children who attended the Denver test examination (i.e., 529 in the treatment group and 544 in the control group). The family survey was conducted for 1,189 children (i.e., 569 in the treatment group and 620 in the control group).[10]

We use the inverse probability weighting method to address missing data problems (Tsiatis, 2006). The IPW estimator solves the following minimization problem:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \left(\frac{s_i}{\hat{p}(\boldsymbol{z_i})} \right) (y_i - \boldsymbol{x_i'\beta})^2$$

where s_i is the indicator if we can use observation i, $P(\boldsymbol{z_i})$ is the propensity score of observation i being observed. $\boldsymbol{z_i}$ is a vector of baseline variables which are always observed for everyone. Here, we mainly focus on the estimation of propensity scores at different stages.

[10]The data missing in the two year annual evaluations can be treated as independent events (e.g., there is no evidence to show being missing from the July 2016 evaluation is correlated with being missing from July 2017 evaluation).

We present the distribution of the propensity score of the samples with missing data in Figure G.2 and compare them with the distribution in the observed outcome samples. In the figures, we show the distribution of the probabilities of being missing at three stages for the observed samples and missing data samples. Our model performs very well in predicting missingness: for observed outcomes samples, the probability of missing is low. For most of them the probability is less than 0.2. For the samples with missing outcome data, the propensity scores are close to being uniformly distributed. For both rounds, the estimated propensity scores are far away from 1 which means we do not need to trim the data. We thus avoid the inconsistency due to data trimming (Maasoumi and Wang, 2019).

In Table G.3, we also report if we only consider the samples for the children who were younger than 2 years old in Sept 2015, we can pass the balance test. We also provide the age distribution for the children who were younger than 2 years old in Sept 2015 in Appendix J.

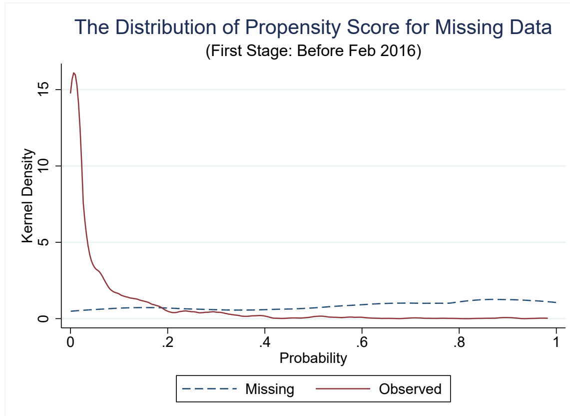
Table G.2: Propensity Score for Missing Data

	First Stage	Second Stage		Third Stage	
	Miss Before January 2016	Miss Denver At Midline	Miss Survey At Midline	Miss Denver At Endline	Miss Survey At Endline
Older Than 24 Months	1.5674***	0.0568	0.0654	0.2484^{+}	0.1042
(In September 2015)	(0.2163)	(0.1491)	(0.1548)	(0.1361)	(0.1300)
Monthly Age	0.0341*	0.0135	0.0151^{+}	-0.0084	0.0012
(In September 2015)	(0.0160)	(0.0087)	(0.0092)	(0.0076)	(0.0072)
The population of the Village	0.0011	-0.0004	-0.0003	0.0003	0.0005^{+}
	(0.0008)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Number of Households in the Village	-0.0001	-0.0000	0.0002	0.0001	0.0004
	(0.0010)	(0.0004)	(0.0005)	(0.0004)	(0.0004)
Size of Working Population	-0.0042**	0.0004	0.0006	-0.0010^{+}	-0.0008
	(0.0016)	(0.0006)	(0.0006)	(0.0005)	(0.0005)
Poor Village or Not	1.3732**	0.1078	0.2037	-0.0161	0.0833
	(0.5066)	(0.2058)	(0.2129)	(0.1833)	(0.1702)
Number of Persons Receiving Social Welfare	-0.0051**	0.0006	-0.0005	0.0025***	0.0017**
	(0.0017)	(0.0008)	(0.0009)	(0.0007)	(0.0007)
Mean Years of Schooling among Villagers	-0.5857*	0.1179	0.1485	-0.1169	-0.1114
	(0.2714)	(0.1230)	(0.1218)	(0.1083)	(0.1020)
Fraction of Interviewed Children Who Are Left-Behind	1.1623	-0.4473	0.7090	-1.0534	-1.3794^{+}
	(2.6825)	(0.9716)	(0.9408)	(0.8324)	(0.7728)
HOME - Sum of Warmth/Responsiveness Items	0.0253	0.0247	0.0153	0.0049	-0.0063
	(0.0410)	(0.0271)	(0.0284)	(0.0239)	(0.0226)
HOME - Sum of Verbal Skills Items	0.1647^{+}	0.0661	0.0992	-0.0273	-0.0643
	(0.0908)	(0.0587)	(0.0626)	(0.0498)	(0.0478)
HOME - Sum of Harshness/Discipline Items	-0.0201	0.0399	0.0448	-0.2120*	-0.1247^{+}
	(0.1206)	(0.0863)	(0.0882)	(0.0876)	(0.0778)
HOME - Sum of Stimulation/Teaching Items	0.0046	-0.0182	-0.0206	-0.0179	-0.0060
	(0.0263)	(0.0181)	(0.0189)	(0.0161)	(0.0154)
HOME - Dum of Outings Items	0.1590**	0.0305	0.0323	-0.0356	-0.0311
	(0.0614)	(0.0430)	(0.0448)	(0.0383)	(0.0362)
Fraction of Children Taking Nutrition Package	-0.3803*	-0.2581*	-0.1977^{+}	-0.0393	-0.1036
	(0.1533)	(0.1011)	(0.1063)	(0.0887)	(0.0842)
Fraction of Children Taking Nutrition Package without Interruption	-0.5557*	-0.0612	-0.0765	0.0259	0.0342
	(0.2332)	(0.1340)	(0.1401)	(0.1144)	(0.1091)
Number of Eligible Kids Living at Home in Interviewed Households in This Village	0.8225***	-0.0958*	-0.0681	-0.0574	-0.0614^{+}
	(0.1450)	(0.0416)	(0.0433)	(0.0353)	(0.0337)
Fraction of Parents Willing to Participate in This Village	0.3958	0.0042	-0.0963	-0.1754	-0.0485
	(0.3759)	(0.2008)	(0.2029)	(0.1720)	(0.1707)
Number of Eligible Kids in Households That Would Be Willing to Participate	-0.7809***	0.0992*	0.0723	0.0592	0.0508
	(0.1464)	(0.0454)	(0.0471)	(0.0381)	(0.0364)
Fraction of Interviewed Households Planning to Migrate with the Child	0.1123	0.1415	0.1828	0.1495	0.0961
	(0.2381)	(0.1565)	(0.1613)	(0.1415)	(0.1374)
Distance between Home Visitor's Home and the Village	0.2559***	0.0126	0.0071	-0.0013	-0.0046
	(0.0666)	(0.0105)	(0.0057)	(0.0063)	(0.0064)
Whether Living in Chengguan Village	0.6431**	-0.6649*	-0.9191**	0.1176	0.1962
	(0.1975)	(0.2681)	(0.3190)	(0.1898)	(0.1809)
Family Migrate out of County	2.5438***	1.6588*	1.7525**		
	(0.3410)	(0.6442)	(0.6755)		
Refuse Home Visit	2.8099***	-5.0985	-5.1383		
	(0.3545)	(167.8907)	(166.4867)		
Refuse Home Visit in September 2015		6.3453	6.3040		
		(167.8891)	(166.4851)		
Constant	-4.1707	-2.2239*	-2.9191**	0.7335	0.9051
	(2.5367)	(1.0230)	(0.9568)	(0.9001)	(0.8452)
Observations	1823	1576	1576	1576	1576

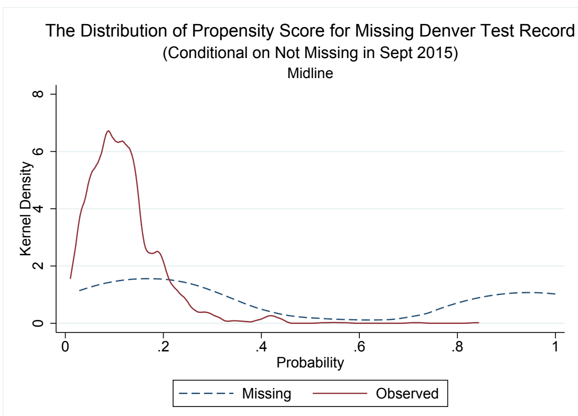
Standard errors are presented in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure G.2: Missing Data Propensity Score Distributions

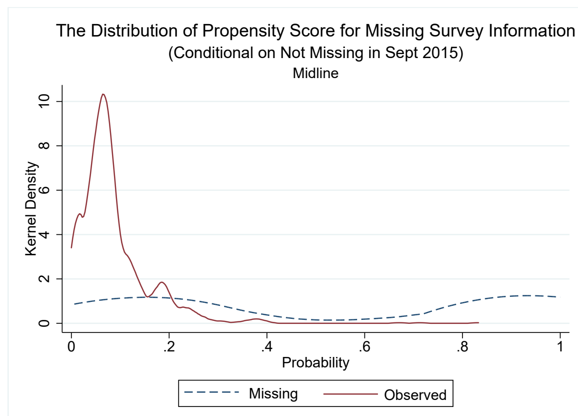
(a)



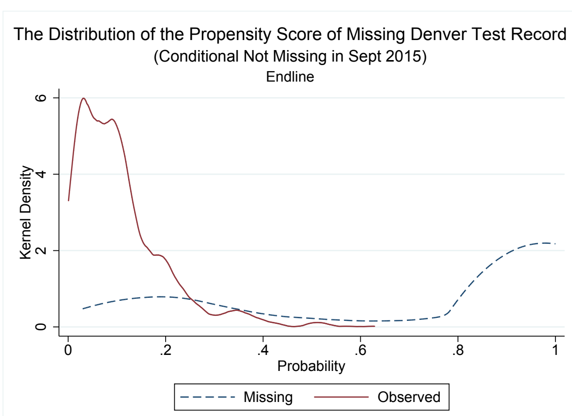
(b)



(c)



(d)



(e)

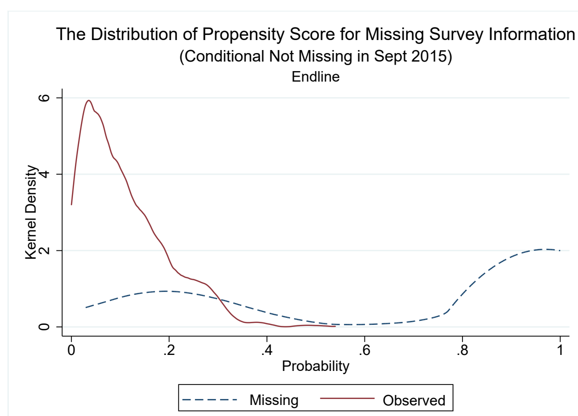


Table G.3: Missing Data Pattern Balance Check

	(1) Missed at Midline	(2) Missed at Endline
Treatment Status	-0.088	-0.119
	(0.177)	(0.077)
Monthly Ages	0.001	-0.015
	(0.869)	(0.019)
Fraction of interviewed children who are left-behind Children	0.179	-0.347
	(0.873)	(0.704)
HOME Warmth	-0.317	-0.065
	(0.092)	(0.662)
HOME Verbal Skills	0.025	-0.009
	(0.781)	(0.862)
HOME Stimulation	-0.167	-0.155
	(0.419)	(0.181)
HOME Outings items	-0.364*	-0.197
	(0.035)	(0.119)
Total HOME score	0.204*	0.088
	(0.031)	(0.381)
Fraction of children taking nutrition package	-0.803	0.459
	(0.162)	(0.392)
Fraction of children taking nutrition package without interruption	0.493	-0.276
	(0.454)	(0.723)
Number of eligible kids living at home in interviewed households in this village	0.053	-0.000
	(0.492)	(0.988)
Fraction of parents willing to participate in this village	0.921	0.349
	(0.208)	(0.619)
Number of eligible kids in households that would be willing to participate	-0.075	0.003
	(0.296)	(0.938)
Fraction of interviewed households planning to migrate with the child	-0.121	-0.398
	(0.850)	(0.562)

Wild bootstrap p values are reported in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

H Estimates with IPW

Table H.1: Treatment Effects on Standardized Denver Scores

	(1)	(2)	(3)	(4)	(5)
	All	All	All	Children $\leq$ 2 Yrs at Enrollment	Children $\leq$ 2 Yrs at Enrollment
			Midline		
Language and Cognitive	0.589***	0.631***	0.714***	0.674***	0.741***
	[0.234, 0.965]	[0.237, 1.036]	[0.319, 1.093]	[0.279, 1.067]	[0.350, 1.144]
Fine Motor	0.334	0.559	0.633*	0.629*	0.703*
	[-0.140, 0.787]	[-0.032, 1.174]	[0.003, 1.313]	[0.023, 1.324]	[0.057, 1.375]
Social-Emotional	0.690**	0.865***	0.879***	0.624***	0.620***
	[0.260, 1.117]	[0.421, 1.312]	[0.467, 1.289]	[0.129, 1.118]	[0.204, 1.067]
Gross Motor	-0.051	-0.004	-0.015	0.054	0.010
	[-0.598, 0.478]	[-0.564, 0.577]	[-0.567, 0.554]	[-0.514, 0.640]	[-0.559, 0.584]
			Endline		
Language and Cognitive	0.979***	0.914***	1.036***	1.016***	1.113***
	[0.585, 1.402]	[0.495, 1.347]	[0.644, 1.458]	[0.637, 1.408]	[0.723, 1.510]
Fine Motor	0.585**	0.574**	0.676***	0.561**	0.645**
	[0.006, 0.956]	[0.067, 1.091]	[0.180, 1.170]	[0.030, 1.095]	[0.139, 1.158]
Social-Emotional	-0.201	-0.276	-0.222	-0.167	-0.115
	[-0.596, 0.202]	[-0.688, 0.123]	[-0.636, 0.194]	[-0.553, 0.215]	[-0.491, 0.275]
Gross Motor	0.067	0.125	0.173	0.155	0.219
	[-0.479, 0.632]	[-0.392, 0.645]	[-0.322, 0.668]	[-0.406, 0.732]	[-0.294, 0.775]
Pre-Treatment Covariates	No	No	Yes	No	Yes
IPW	No	Yes	Yes	Yes	Yes

Notes: For a more comprehensive set of estimates not adjusting for IPW and not adjusting for pre-treatment covariates. See Appendix **?**.

1. The 95% confidence intervals in brackets are constructed using the wild bootstrap clustered at the village level.

2. The mean and variance for the standardized score are estimated from the pooled sample of the control group children.

3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4. The negative treatment effects for social-emotional ability vanish after we adjust for item difficulty.

5. The columns with the label “All” include all the observations, and the columns with the label “Children $\leq$ 2 Yrs at Enrollment” restrict the sample to the children who were under 2 years old when they enrolled in the program.

Table H.2: Treatment Effects on Standardized Denver Scores

(Female)

	(1)	(2)	(3)	(4)	(5)
	All	All	All	Children ≤ 2 Yrs at Enrollment	Children ≤ 2 Yrs at Enrollment
			Midline		
Language and Cognitive	0.410	0.417	0.445	0.511**	0.534**
	[-0.076, 0.869]	[-0.035, 0.884]	[-0.014, 0.910]	[0.040, 0.991]	[0.080, 0.990]
Fine Motor	0.400	0.399	0.335	0.512	0.544
	[-0.252, 1.049]	[-0.271, 1.065]	[-0.269, 1.211]	[-0.088, 1.142]	[-0.082, 1.189]
Social-Emotional	1.020***	1.068***	1.114***	0.912**	0.938***
	[0.445, 1.614]	[0.520, 1.614]	[0.681, 1.550]	[0.272, 1.541]	[0.400, 1.431]
Gross Motor	0.117	0.063	0.058	0.085	0.019
	[-0.487, 0.751]	[-0.565, 0.665]	[-0.532, 0.675]	[-0.514, 0.725]	[-0.605, 0.652]
			Endline		
Language and Cognitive	0.852**	0.895**	0.950**	0.865**	0.893**
	[0.077, 1.596]	[0.159, 1.612]	[0.213, 1.675]	[0.122, 1.590]	[0.177, 1.598]
Fine Motor	0.804**	0.815**	0.866**	0.836**	0.855**
	[0.111, 1.500]	[0.088, 1.553]	[0.189, 1.574]	[0.110, 1.554]	[0.117, 1.579]
Social-Emotional	-0.264	-0.298	-0.309	-0.264	-0.291
	[-0.806, 0.254]	[-0.805, 0.267]	[-0.775, 0.160]	[-0.859, 0.342]	[-0.820, 0.206]
Gross Motor	0.188	0.246	0.257	0.460	0.445
	[-0.737, 1.091]	[-0.668, 1.094]	[-0.582, 1.080]	[-0.410, 1.308]	[-0.417, 1.326]
Pre-Treatment Covariates	No	No	Yes	No	Yes
IPW	No	Yes	Yes	Yes	Yes

Notes: 1. The 95% confidence intervals in brackets are constructed using the wild bootstrap clustered at the village level.

2. The mean and variance for the standardized score are estimated from the pooled sample of the control group children.

3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4. The negative treatment effects for social-emotional ability vanish after we adjust for item difficulty.

5. The columns with the label "All" include all the observations, and the columns with the label "Children ≤ 2 Yrs at Enrollment" restrict the sample to the children who were under 2 years old when they enrolled in the program.

Table H.3: Treatment Effects on Standardized Denver Scores

(Male)

	(1)	(2)	(3)	(4)	(5)
	All	All	All	Children $\leq$ 2 Yrs at Enrollment	Children $\leq$ 2 Yrs at Enrollment
			Midline		
Language and Cognitive	0.747***	0.852***	0.938***	0.896***	0.911***
	[0.236, 1.257]	[0.261, 1.462]	[0.389, 1.499]	[0.345, 1.460]	[0.329, 1.501]
Fine Motor	0.395	0.674	0.716	0.730	0.771
	[-0.108, 0.908]	[-0.083, 1.532]	[-0.099, 1.598]	[-0.028, 1.577]	[-0.070, 1.747]
Social-Emotional	0.436	0.589*	0.549**	0.395	0.280
	[-0.115, 0.989]	[0.028, 1.140]	[0.047, 1.054]	[-0.178, 0.946]	[-0.272, 0.842]
Gross Motor	-0.066	0.079	-0.041	0.152	-0.021
	[-0.798, 0.661]	[-0.728, 0.900]	[-0.700, 0.639]	[-0.634, 0.963]	[-0.682, 0.659]
			Endline		
Language and Cognitive	1.050***	0.797**	0.950***	1.000***	1.111***
	[0.514, 1.560]	[0.205, 1.436]	[0.448, 1.497]	[0.468, 1.513]	[0.625, 1.626]
Fine Motor	0.460	0.388	0.462	0.346	0.388
	[-0.212, 1.117]	[-0.314, 1.108]	[-0.206, 1.144]	[-0.374, 1.042]	[-0.355, 1.124]
Social-Emotional	-0.139	-0.306	-0.256	-0.157	-0.169
	[-0.643, 0.390]	[-0.895, 0.305]	[-0.829, 0.326]	[-0.654, 0.351]	[-0.701, 0.400]
Gross Motor	-0.059	-0.071	-0.048	-0.169	-0.138
	[-0.528, 0.424]	[-0.543, 0.407]	[-0.510, 0.419]	[-0.663, 0.332]	[-0.629, 0.359]
Pre-Treatment Covariates	No	No	Yes	No	Yes
IPW	No	Yes	Yes	Yes	Yes

Notes: 1. The 95% confidence intervals in brackets are constructed using the wild bootstrap clustered at the village level.
2. The mean and variance for the standardized score are estimated from the pooled sample of the control group children.
3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
4. The negative treatment effects for social-emotional skills vanish after we adjust for item difficulty.
5. The columns with the label "All" include all the observations, and the columns with the label "Children $\leq$ 2 Yrs at Enrollment" restrict the sample to the children who were under 2 years old when they enrolled in the program.

I Estimates Using Matching

We examine the robustness of our estimates and estimate treatment effects from matching estimation method. The results are very close to OLS regression results. We use two different matching methods: one based on Mahalanobis weights and one based on propensity scores. For each method, we try three different specifications: one match per observation, two matches per observation, and three matches per observation.

Table I.1: Treatment Effects from Nearest Neighbor Matching

	Standardized Denver Scores			
	Language and Cognitive	Social-Emotional	Fine Motor	Gross Motor
		Midline		
One Match per Observation	0.685***	0.785***	0.202	-0.161
	(0.203)	(0.238)	(0.229)	(0.195)
Two Matches per Observation	0.71***	0.776***	0.286	-0.11
	(0.188)	(0.224)	(0.214)	(0.182)
Three Matches per Observation	0.701***	0.762***	0.253*	-0.069
	(0.178)	(0.213)	(0.202)	(0.184)
		Endline		
One Match per Observation	0.583*	-0.528**	0.373	-0.001
	(0.257)	(0.201)	(0.224)	(0.272)
Two Matches per Observation	0.772***	-0.427*	0.482*	-0.086
	(0.238)	(0.187)	(0.206)	(0.235)
Three Matches per Observation	0.865***	-0.392*	0.504*	-0.011
	(0.227)	(0.182)	(0.198)	(0.227)

Notes: 1. Standard errors are constructed using results in Abadie and Imbens (2012).
2. Matching covariates include monthly ages, gender, the language skill scores on the HOME IT scale, the learning materials score on the HOME IT scale, the take-up rate of a nutrition supplement program in the village, the compliance rate for a countywide nutrition program in the village, the percentage of left-behind children in the children sample, the per capita net income in the village, the average years of schooling in the village, the percentage of caregivers intending to participate in the parenting intervention program, and the percentage of families intending to bring the child when migrating to urban areas.
3. Matching matrix is Mahalanobis matrix.
4. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

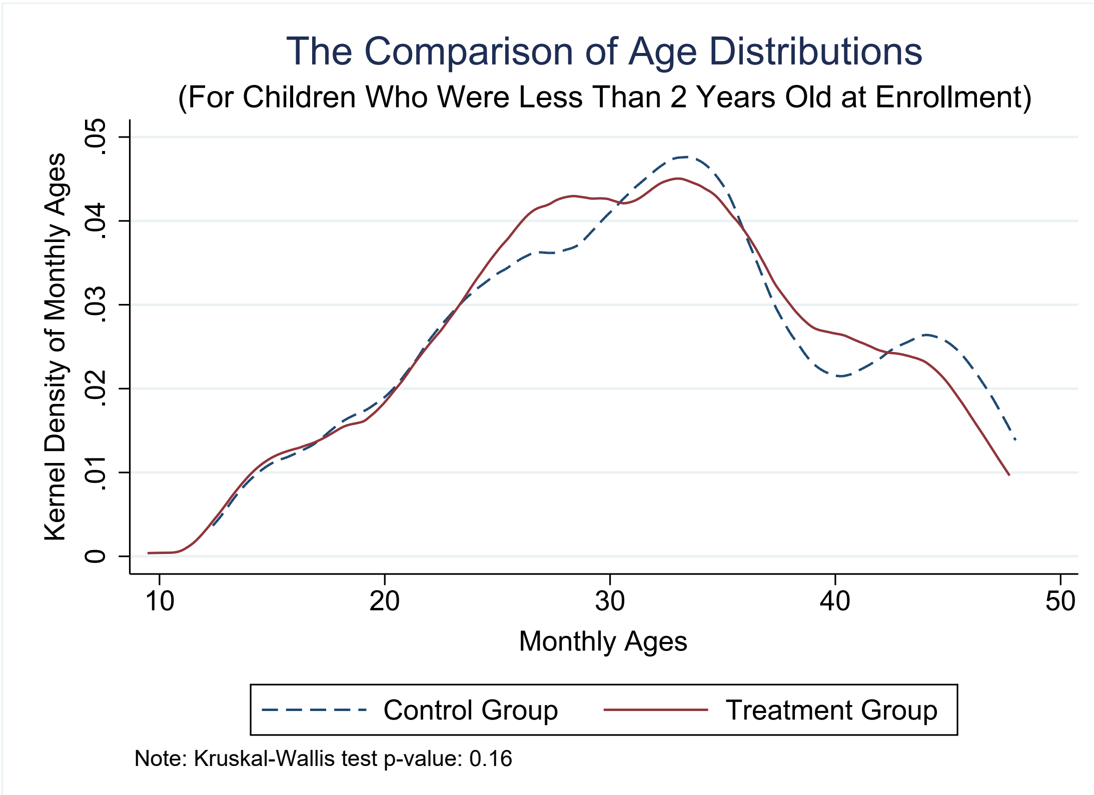
Table I.2: Treatment Effects from Propensity Score Matching

	Standardized Denver Scores			
	Language and Cognitive	Social-Emotional	Fine Motor	Gross Motor
		Midline		
One Match per Observation	0.858***	0.663**	0.284	0.140
	(0.187)	(0.224)	(0.211)	(0.194)
Two Matches per Observation	0.759***	0.745***	0.268	0.185
	(0.172)	(0.209)	(0.202)	(0.195)
Three Matches per Observation	0.765***	0.740***	0.270	0.183
	(0.165)	(0.210)	(0.203)	(0.189)
		Endline		
One Match per Observation	1.033***	-0.257	0.743***	0.005
	(0.269)	(0.203)	(0.221)	(0.227)
Two Matches per Observation	1.072***	-0.176	0.796***	0.103
	(0.242)	(0.198)	(0.212)	(0.207)
Three Matches per Observation	1.018***	-0.217	0.695***	0.106
	(0.219)	(0.195)	(0.202)	(0.199)

Notes: 1. Standard errors are robust and constructed by Abadie and Imbens (2012).
2. Matching covariates include monthly ages, gender, the language skill scores on the HOME IT scale, the learning materials score on the HOME IT scale, the take-up rate of a nutrition supplement program in the village, the compliance rate for a countywide nutrition program in the village, the percentage of left-behind children in the children sample, the per capita net income in the village, the average years of schooling in the village, the percentage of caregivers intending to participate in the parenting intervention program, and the percentage of families intending to bring the child when migrating to urban areas.
3. Propensity score is estimated by probit model.
4. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

J Age Balance Test

Figure J.1: Comparison of Monthly Ages by Treatment Status



This figure plots the age of children who were less than two years old at enrollment. In Appendix G, we document that, for some reason, the children who were older than two years old in the treatment group were not enrolled in the treatment group, which causes an imbalance.

Figure J.1 shows that the age distribution is balanced and that we cannot reject the null that distributions are the same.

We add the baseline comparison for the children enrolled after the experiment initiated and find that the baselines are very close between the treatment and control groups for these new enrolled children.

K Normalization Method in Anderson and Rubin (1956)

Factor models are notorious for being identified through arbitrary assumptions about how the factors enter different equations. This led to their disuse after their introduction into economics by Jöreskog and Goldberger (1975), Goldberger (1972), Chamberlain and Griliches

(1975), and Chamberlain (1977a,b). The essential identification problem in factor analysis is clearly stated by Anderson and Rubin (1956). Williams (2020) has an illuminating generalization of their conditions. If there are L measurements on K mutually independent factors arrayed in a vector θ, we may write outcomes G in terms of latent variables θ as

$$G = \mu + \Lambda\theta + \varepsilon, \tag{1}$$

where G is $L \times 1$, $\theta \perp\!\!\!\perp \varepsilon$, μ is an $L \times 1$ vector of means, which may depend on X, θ is $K \times 1$, ε is $L \times 1$ and $\mathbf{\Lambda}$ is $L \times K$. $\varepsilon_i \perp\!\!\!\perp \varepsilon_j, i, j = 1, \ldots, L, i \neq j$. Even if $\theta_i \perp\!\!\!\perp \theta_j$, $i \neq j, i, j = 1, \ldots, K$, the model is underidentified. System (1) characterizes the factor structure for both treatment and control groups but values of the variables and parameters may vary across treatment groups.

Using only the information in the covariance matrices, as is common in factor analysis,

$$\operatorname{cov}(G) = \Lambda\Sigma_\theta\Lambda' + D_\varepsilon \tag{2}$$

where Σ_θ is a diagonal matrix of the variances of the factors, and D_ε is a diagonal matrix of the "uniqueness" variances. We observe G but not θ or ε, and we seek to identify Λ, Σ_θ, and D_ε. Without some restrictions, this is clearly an impossible task. Conventional factor-analytic models make assumptions to identify parameters. The restriction that the components of θ are independent is one restriction that we have already made, but it is not enough. The diagonals of $\operatorname{cov}(G)$ combine elements of D_ε with parameters from the rest of the model. Once those other parameters are determined, the diagonals identify D_ε. Accordingly, we can only rely on the $\frac{L(L-1)}{2}$ non-diagonal elements to identify the K variances (assuming $\theta_i \perp\!\!\!\perp \theta_j,\ \forall i \neq j$), and the $L \times K$ factor loadings. Since the scale of each θ_i is arbitrary, one factor loading devoted to each factor is normalized to unity to set the scale. Accordingly, we require that

$$\underbrace{\frac{L(L-1)}{2}}_{\text{Number of off-diagonal covariance elements}} \geq \underbrace{(L \times K - K)}_{\text{Number of unrestricted } \Lambda} + \underbrace{K}_{\text{Variances of } \theta}$$

so

$$L \geq 2K + 1$$

is a necessary condition for identification. These are sometimes called Lederman bounds.

The strategy pursued in this paper is transparent and assumes that there are two or more elements of G devoted *exclusively* to factor θ_1, and at least three elements of G that are generated by factor θ_1, two or more other elements of G devoted only to factors θ_1 and θ_2, with at least three elements of G that depend on θ_1 and θ_2, and so forth.A subset of test scores may only proxy cognitive ability. Other measurements may proxy only soft skills. In general, however, multiple abilities generate behavior and measurement. "Dedicated measurements" are those with factor loadings only in one column. This assumption is commonly used in the literature on child development.

Order G under this assumption so that we get the following pattern for Λ (we assume that the displayed λ_{ij} are not zero):

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 & \vdots & ... & ... & 0 \\ \lambda_{21} & 0 & 0 & 0 & \vdots & ... & ... & 0 \\ \lambda_{31} & 1 & 0 & 0 & \vdots & ... & ... & 0 \\ \lambda_{41} & \lambda_{42} & 0 & 0 & \vdots & ... & ... & 0 \\ \lambda_{51} & \lambda_{52} & 1 & 0 & \vdots & ... & ... & 0 \\ \lambda_{61} & \lambda_{62} & \lambda_{63} & 0 & \vdots & ... & ... & 0 \\ \lambda_{71} & \lambda_{72} & \lambda_{73} & 1 & \vdots & 0 & ... & 0 \\ \lambda_{81} & \lambda_{82} & \lambda_{83} & \lambda_{84} & \vdots & 0 & ... & 0 \\ ... & ... & ... & ... & \vdots & ... & ... & ... \\ \lambda_{L,1} & \lambda_{L,2} & \lambda_{L,3} & ... & \vdots & ... & ... & \lambda_{L,K} \end{pmatrix}. \tag{3}$$

Assuming nonzero covariances

$$\text{cov}(g_j, g_l) = \lambda_{j1}\lambda_{l1}\sigma^2_{\theta_1}, \quad l = 1, 2; \quad j = 1, ..., L; \quad j \neq l.$$

where

$$\lambda_{11} = 1, \lambda_{3,2} = 1, \lambda_{5,3} = 1, ... \quad . \tag{4}$$

In particular

$$\begin{aligned} \text{cov}(g_1, g_\ell) &= \lambda_{\ell 1}\sigma^2_{\theta_1} \\ \text{cov}(g_2, g_\ell) &= \lambda_{\ell 1}\lambda_{21}\sigma^2_{\theta_1}. \end{aligned}$$

Assuming $\lambda_{\ell 1} \neq 0$, we obtain

$$\frac{\text{cov}(g_2, g_\ell)}{\text{cov}(g_1, g_\ell)} = \lambda_{21}.$$

Hence, from $\text{cov}(g_1, g_2) = \lambda_{21}\sigma^2_{\theta_1}$, we obtain $\sigma^2_{\theta_1}$, and hence $\lambda_{\ell 1}$, $\ell = 1, \ldots, L$. We can proceed to the next set of two measurements and identify

$$\text{cov}(g_l, g_j) = \lambda_{l1}\lambda_{j1}\sigma^2_{\theta_1} + \lambda_{l2}\lambda_{j2}\sigma^2_{\theta_2}, \quad l = 3, 4; \quad j \geq 3; \quad j \neq l.$$

Since we know the first term on the right hand side by the previous argument, we can proceed using $\text{cov}(g_l, g_j) - \lambda_{l1}\lambda_{j1}\sigma^2_{\theta_1}$ and identify the $\lambda_{j2}, j = 1, \ldots, L$ using the previous line of reasoning (some of these elements are fixed to zero). Proceeding in this fashion, we can identify Λ and Σ_θ subject to diagonal normalizations. This argument works for all but the system for the Kth and final factor. Observe that for all of the preceding factors there are at least three measurements that depend on $\theta_j, j = 1, \ldots, K-1$, although only two of the measurements need to depend solely on $\theta_1, \ldots, \theta_{K-1}$, respectively. To obtain the necessary three measurements for the Kth and final factor, we require that there be at least three outcomes with measurements that depend on $\theta_1, \ldots, \theta_K$.

Knowing Λ and Σ_θ, we can identify D_ε. Use of partially dedicated measurement systems for specific factors and panel data helps to eliminate much of the arbitrariness that plagued factor analysis during its introduction in economics in the 1970s. Although many other restrictions on the model are possible, the one we adopt has the advantage of simplicity and interpretability in many contexts. Williams (2020) offers other possible normalizations.

L Robustness Checks for Factor Normalization

Factor models require normalizations as long as we seek to separate factors from factor loadings. We use the items in the category "Baseline" to normalize the first four loadings for the four skills studied in this paper. We use a self-explanatory simplified notation. The items listed assign $\alpha^{jk} = 1$ to the items listed and zero otherwise, while the remaining factor loadings are freely specified. In Table L.2, we show means of the latent factor loadings under different normalizations and also test whether they are different from our original normalization estimates. We find that the results are quite stable across different normalization choices if we choose the normalized items in the medium difficulty level range.

Table L.1: The List of Normalized Task Items

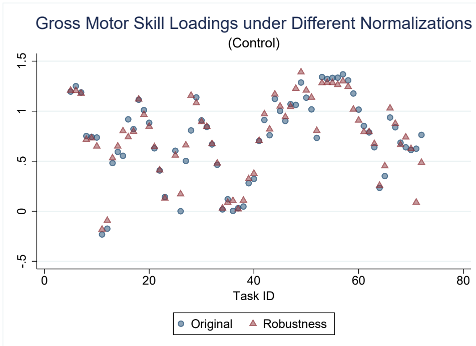
	Original	(1)	(2)	(3)	(4)	(5)
Social-Emotional	Wash and Dry Hands	Wash and Dry Hands	Wash and Dry Hands	Wash and Dry Hands	Drink From Cup	Drink From Cup
Fine Motor	Imitate Vertical Line	Thumb Wiggle	Imitate Vertical Line	Imitate Vertical Line	Imitate Vertical Line	Thumb Wiggle
Language and Cognitive	Combine Words	Combine Words	Name Body Parts 6	Name Body Parts 6	Name Body Parts 6	Name Body Parts 6
Gross Motor	Broad Jump	Broad Jump	Broad Jump	Balance Each Foot 1 Second	Balance Each Foot 1 Second	Balance Each Foot 1 Second

Table L.2: Skill Loading Means Comparison under Different Normalizations

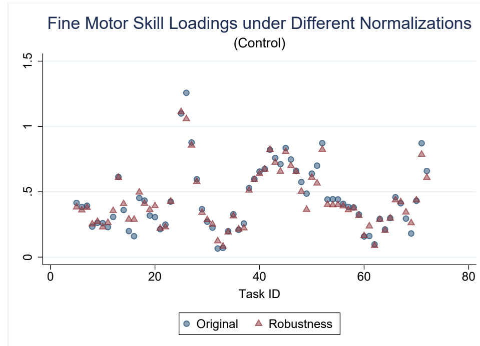
Control							Treatment						
		Different Normalizations							Different Normalizations				
	Original	(1)	(2)	(3)	(4)	(5)		Original	(1)	(2)	(3)	(4)	(5)
	Language and Cognitive Skill Loadings							Language and Cognitive Skill Loadings					
Mean	0.442	0.453	0.442	0.462	0.439	0.454	Mean	0.656	0.697	0.697	0.725	0.721	0.709
S.D.	(0.371)	(0.320)	(0.346)	(0.354)	(0.380)	(0.381)	S.D.	(0.478)	(0.447)	(0.457)	(0.447)	(0.479)	(0.462)
p-value		0.335	0.969	0.125	0.876	0.599	*p*-value		0.008	0.001	0.000	0.037	0.076
	Gross Motor Skill Loadings							Gross Motor Skill Loadings					
Mean	0.712	0.710	0.716	0.741	0.737	0.768	Mean	0.668	0.639	0.690	0.690	0.681	0.692
S.D.	(0.422)	(0.425)	(0.388)	(0.373)	(0.419)	(0.451)	S.D.	(0.453)	(0.405)	(0.433)	(0.397)	(0.418)	(0.416)
p-value		0.897	0.690	0.088	0.345	0.069	*p*-value		0.092	0.013	0.120	0.529	0.299
	Fine Motor Skill Loadings							Fine Motor Skill Loadings					
Mean	0.437	0.431	0.465	0.429	0.451	0.470	Mean	0.539	0.543	0.565	0.556	0.539	0.541
S.D.	(0.269)	(0.248)	(0.263)	(0.249)	(0.300)	(0.292)	S.D.	(0.240)	(0.228)	(0.237)	(0.219)	(0.240)	(0.252)
p-value		0.348	0.005	0.606	0.438	0.082	*p*-value		0.672	0.097	0.390	0.989	0.944
	Social-Emotional Skill Loadings							Social-Emotional Skill Loadings					
Mean	0.259	0.245	0.238	0.227	0.220	0.214	Mean	0.223	0.177	0.190	0.188	0.168	0.171
S.D.	(0.275)	(0.279)	(0.270)	(0.278)	(0.314)	(0.338)	S.D.	(0.260)	(0.227)	(0.227)	(0.202)	(0.276)	(0.290)
p-value		0.119	0.005	0.059	0.138	0.118	*p*-value		0.000	0.001	0.020	0.037	0.046

Figure L.1: The Comparison of Latent Skill Loadings under Different Normalizations

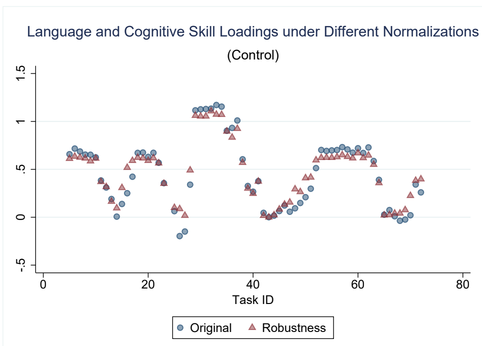
(a)



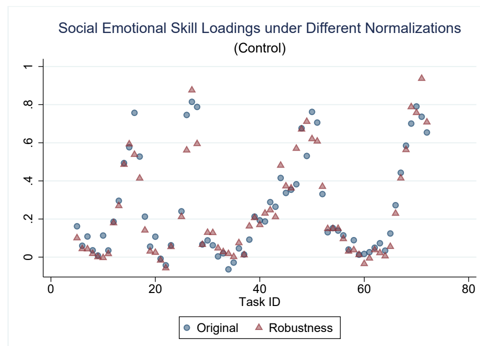
(b)



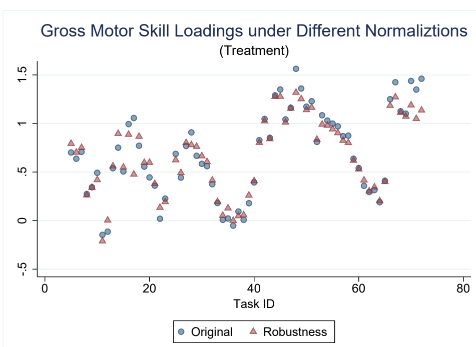
(c)



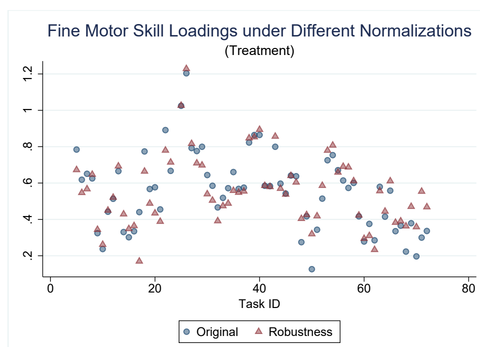
(d)



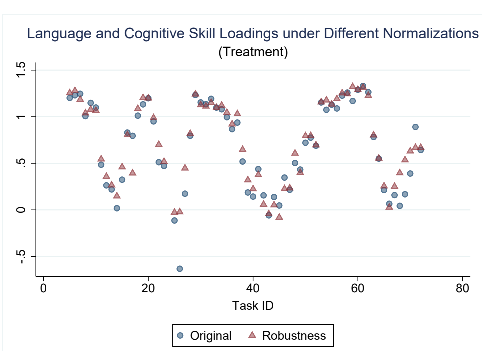
(e)



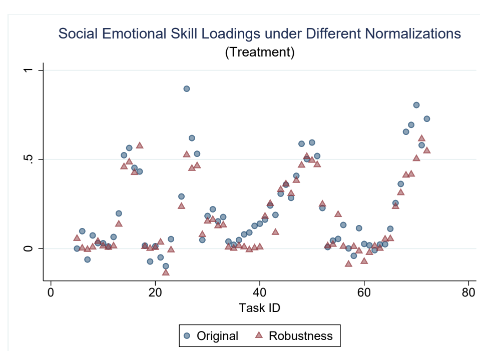
(f)



(g)



(h)



M Consistency and Asymptotic Normality of Individual Factors (Wang, 2020) and a Factor Estimation Procedure from Chen et al. (2021)

We use the analysis of Wang (2020), who proves the consistency and asymptotic normality of estimators of $\boldsymbol{\theta_i^{j(k)}}$ under conditions we satisfy. We actually know some of the factors he estimates, so our model is a special case of his. We apply an estimation procedure for factors proposed by Chen et al. (2021).[11] We use a simplified self-explanatory notation in this appendix to facilitate exposition.

(1) For each iteration k, given the set of parameters $\{\boldsymbol{\beta^{j(k)}}, \boldsymbol{\theta_i^{j(k)}}, \boldsymbol{\alpha^{j(k)}}, \boldsymbol{\delta^{j(k)}}\}$, define $\mu_{ij}^j = \boldsymbol{X^{j\prime}\beta^{(k)}} + \delta^{j(k)} + (\boldsymbol{\theta_i^{j(k)\prime}})\boldsymbol{\alpha^{j(k)}}$.

(2) **E-step**: Calculate

$$\begin{aligned}\hat{Y}_i^{j(k)} &= E(Y_{ij}^{j*}|Y_i^j, \boldsymbol{X_i^j}, \boldsymbol{\beta^{j(k)}}, \boldsymbol{\alpha^{j(k)}}, \delta^{j(k)}, \boldsymbol{\theta_i^{j(k)}}) \\ &= \mu_i^{j(k)} + (Y_i^j - \Phi^{j(k)}(\mu_i^{j(k)}))\phi^{j(k)}(\mu_i)/\{\Phi(\mu_i^{j(k)})(1 - \Phi(\mu_i^{j(k)}))\}.\end{aligned}$$

(3) **M-step** conditional maximization steps:

- Update $\boldsymbol{\beta^j}$: $\boldsymbol{\beta^{j(k+1)}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\hat{Y}^{j(k)} - \delta^{j(k)} - (\boldsymbol{\theta^{j(k)}})'\boldsymbol{\alpha^{j(k)}})$.
- Update δ^j: $\delta^{j(k+1)} = \sum_i(\hat{Y}_i^{j(k)} - \boldsymbol{X_i'\beta^{j(k+1)}} - (\boldsymbol{\theta_i^{j(k)}})'\boldsymbol{\alpha^{j(k)}})/N_{Ij}$.
- Update $\alpha_{j,m}^{j(k+1)}$, where m indicates the mth latent factor, since for each latent factor, we have one item m^* with loading $\alpha_{m^*,m^*} = 1$ and $\alpha_{m^*,j\neq m^*} = 0$

$$\boldsymbol{\alpha}_{j,m}^{(k+1)} = \frac{\sum_i(\hat{Y}_i^{j(k)} - \boldsymbol{X_i^{j\prime}\beta^{j(k+1)}} - \delta^{j(k+1)})(\hat{Y}_i^{m^*(k)} - (\boldsymbol{X_i^{m^*}})'\boldsymbol{\beta^{(k+1)}} - \delta^{m^*(k+1)})}{\sum_i(\hat{Y}_i^{m^*(k)} - (\boldsymbol{X_i^{m^*}})'\boldsymbol{\beta^{j(k+1)}} - \delta^{m^*(k+1)})^2}.$$

- Update $\theta_i^{(k+1)}$, in this step, we use the closed form solution to update the individual level latent factors, which is more robust than the method proposed in Chen et al. (2021).

$$\theta_i^{(k+1)} = (\hat{Y}_i^{j(k)} - \boldsymbol{X_i^{j\prime}\beta^{j(k+1)}} - \delta^{j(k+1)})'\boldsymbol{\alpha^{j(k+1)}}(\boldsymbol{\alpha^{j(k+1)}\alpha'^{j(k+1)}})^{-1}.$$

(4) Iterate until convergence.

[11] Wang's analysis assumes no $\boldsymbol{X_i^j}$ and identifies and develops a consistent estimator of $\boldsymbol{\theta_i^j}$, as well as of factor loadings $\boldsymbol{\alpha^j}$. It is trivial to apply his analysis when components of $\boldsymbol{\theta_i^j}$ are known, i.e., the $\boldsymbol{X_i^j}$.

N Estimates for the Factor Model

Table N.1: Estimated Coefficients for the Observed Covariates

	Control Group	Treatment Group
Monthly Age	0.961	0.924
	[0.166, 1.987]	[0.161, 1.738]
Monthly Age^2	-0.009	-0.009
	[-0.025, 0.002]	[-0.0193, 0.002]
Male	0.356	-0.144
	[-1.081, 2.363]	[-1.178, 1.148]
Constant	-16.756	-15.571
	[-35.260, -2.727]	[-31.620, -2.457]
	$\chi^2(4) = 0.004$	$p = 0.999$

Notes: 1. The values presented in the brackets are 95% confidence intervals.

2. The confidence intervals are calculated by the paired cluster bootstrap at the village level.

3. We use the χ^2 test to examine whether the coefficients of two groups are the same or not. The test results show that we cannot reject the hypothesis that these coefficients are the same.

O Fit of Estimated Models to Sample Data

O.1 Full Model

Figure O.1: Model Fit for Language and Cognitive Tasks

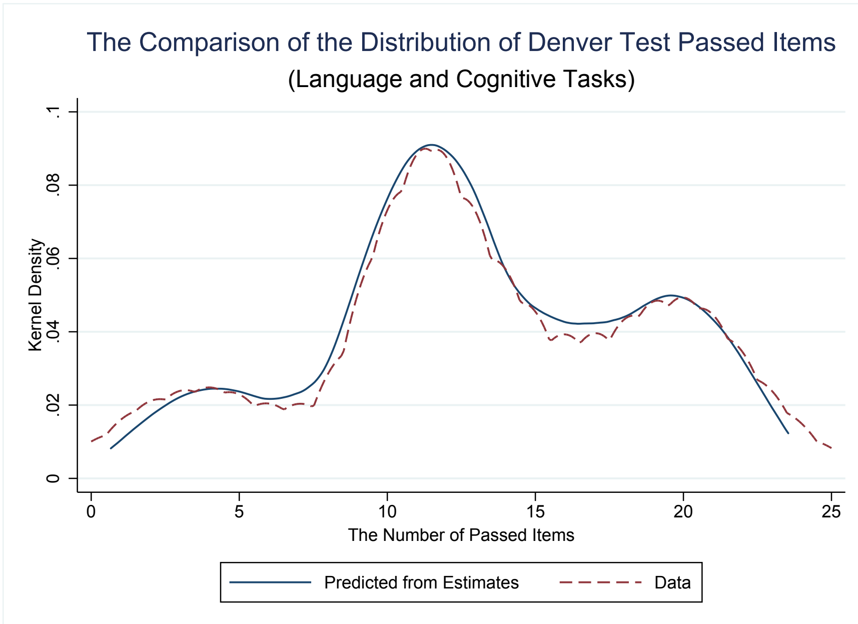


Figure O.2: Model Fit for Social-Emotional Tasks

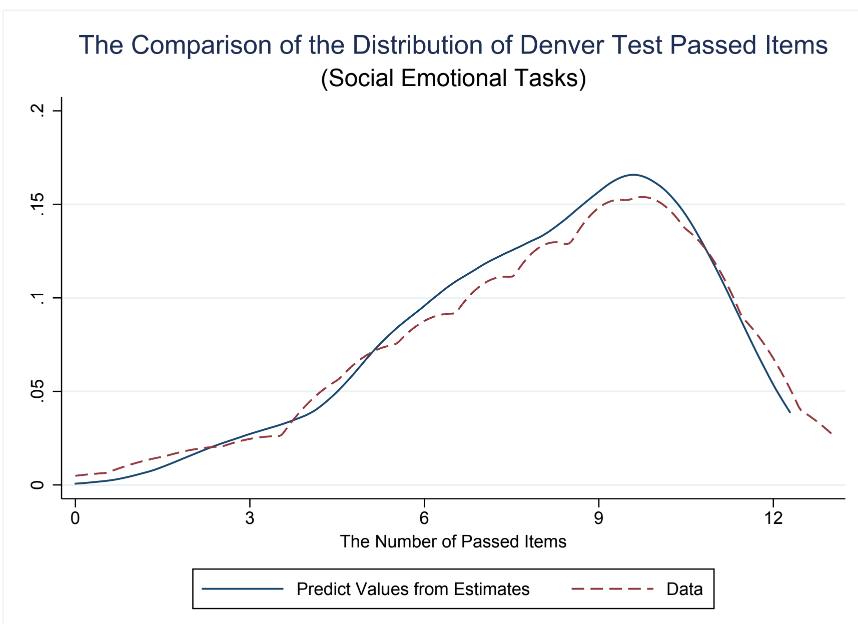


Figure O.3: Model Fit for Fine Motor Tasks

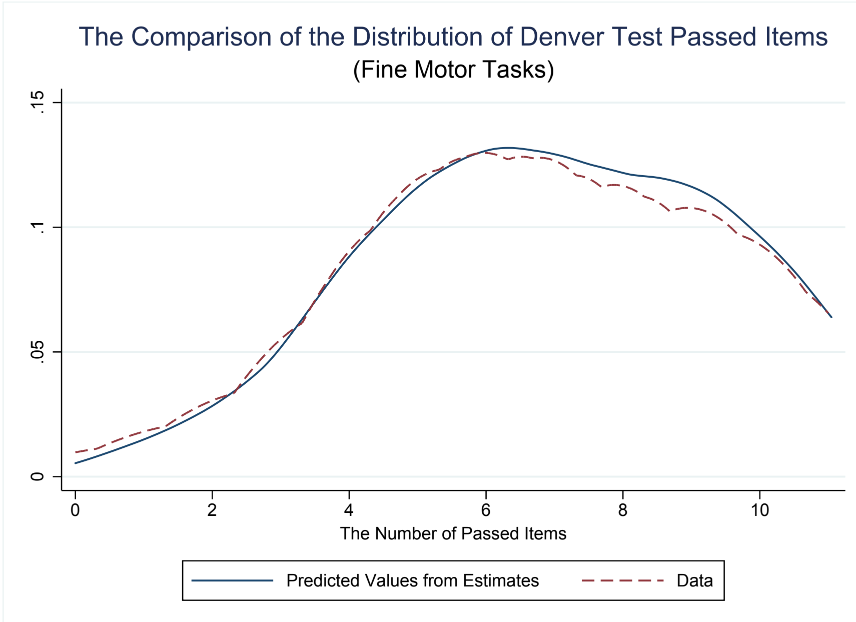
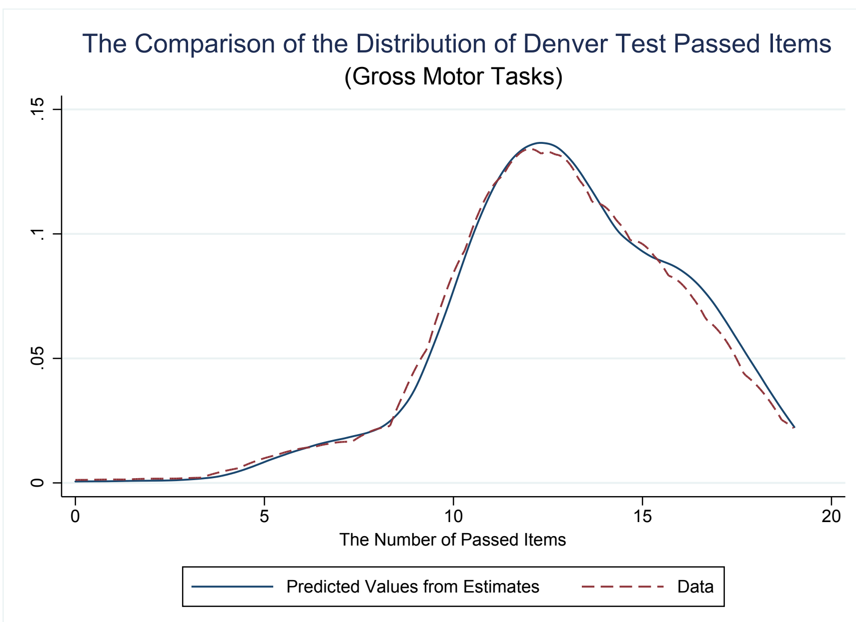


Figure O.4: Model Fit for Gross Motor Tasks



O.2 Restricted Model (without Task Difficulty Parameters)

Figure O.5: Model Fit for Language and Cognitive Tasks

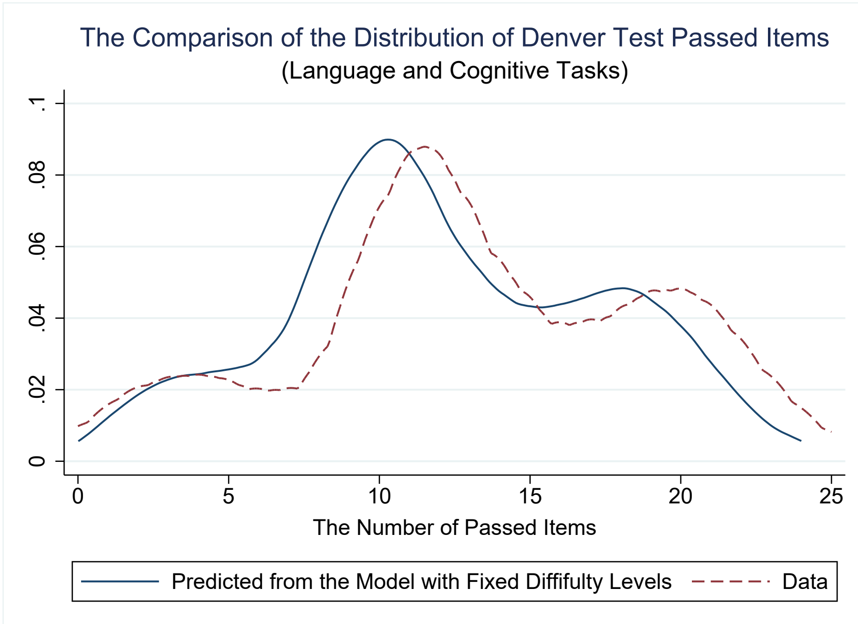


Figure O.6: Model Fit for Social-Emotional Tasks

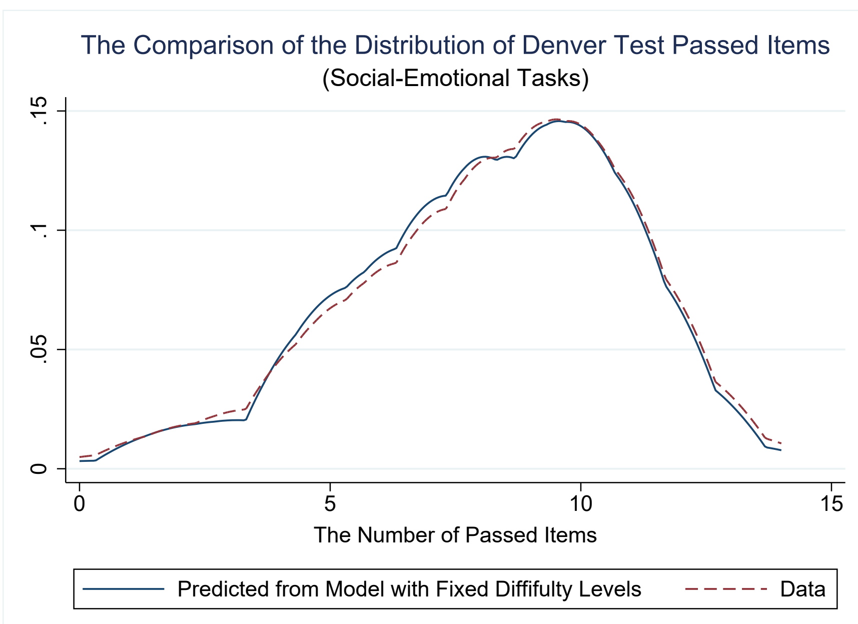


Figure O.7: Model Fit for Fine Motor Tasks

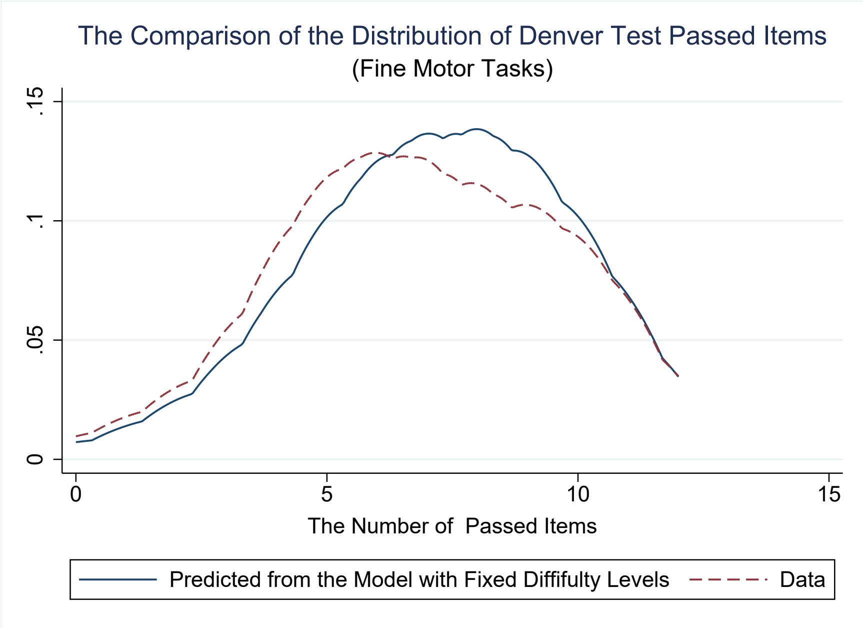
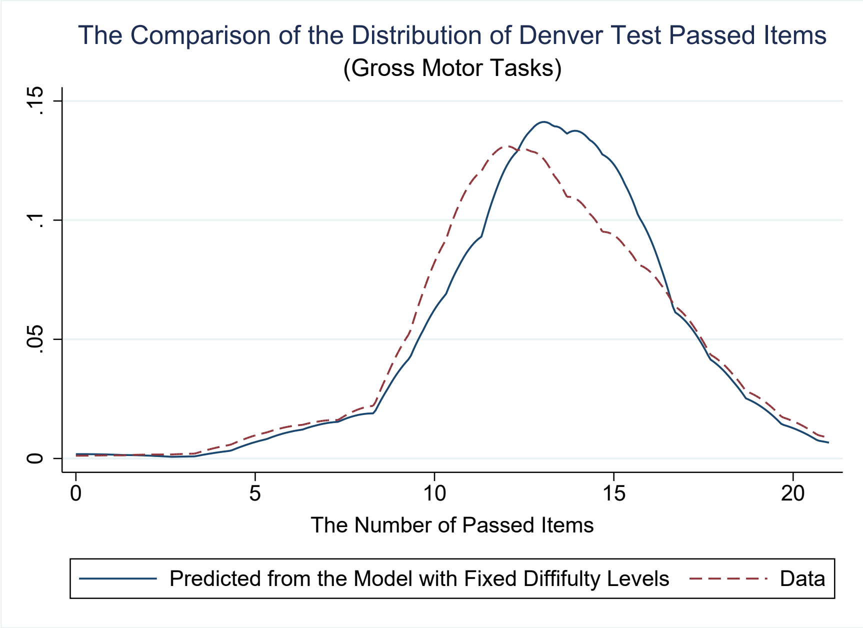


Figure O.8: Model Fit for Gross Motor Tasks



P Robustness of Estimates

In this section, we also estimate the model described in Section 3.3 and report the estimates once we control for more covariates: years of education and HOME environment scores at the baseline in Table P.1.

Table P.1: Estimated Coefficients for the Observed Covariates

	Control Group	Treatment Group
Monthly Age	0.975	0.975
	[0.405, 1.676]	[0.232, 1.989]
Monthly Age^2	-0.008	-0.009
	[-0.184, -0.003]	[-0.203, -0.001]
Male	0.298	0.698
	[-0.676, 1.823]	[-0.529, 1.648]
Father's Years of Education	0.009	0.017
	[-0.335, 0.575]	[-0.178, 0.581]
Mother's Years of Education	0.012	-0.003
	[-0.259, 0.326]	[-0.317, 0.222]
Grandmother's Years of Education	0.003	0.003
	[-0.327, 0.233]	[-0.280, 0.331]
Home: warmth	-0.016	0.012
	[-0.427, 1.012]	[-0.639, 0.572]
Home: verbal skills	0.040	0.060
	[-0.639, 1.278]	[-0.436, 1.830]
Home: hostility	-0.028	0.053
	[-0.636, 1.486]	[-0.198, 1.521]
Home: learning literacy	0.016	0.009
	[-0.347, 0.564]	[-0.399, 0.202]
Home: outings	-0.016	0.012
	[-0.459, 1.428]	[-0.466, 1.294]
Constant	-17.874	-17.000
	[-19.829, -3.909]	[-20.193, -1.126]
	$\chi^2(12) = 0.256$	$p = 0.999$

Notes: 1. The values presented in the brackets are 95% confidence intervals.

2. The confidence intervals are calculated by the paired cluster bootstrap at the village level.

3. We use the χ^2 test to examine whether the coefficients of two groups are the same or not. The test results show that we cannot reject the hypothesis that these coefficients are the same.

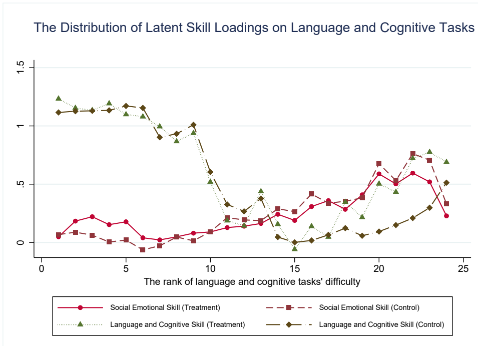
Q Estimates of Skill Loadings

Q.1 Estimates of Skill Loadings

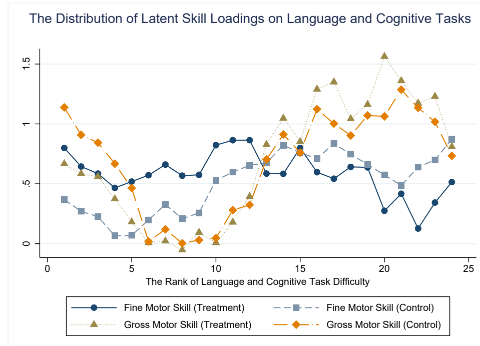
We plot the point estimates of loadings for each task in Figure Q.1. We find that the scale of language and cognitive skill loadings on language and cognitive tasks is much higher than the loadings of social-emotional skill on language and cognitive tasks. This finding also holds with other types of tasks.

Figure Q.1: The Distribution of Latent Skill Loadings

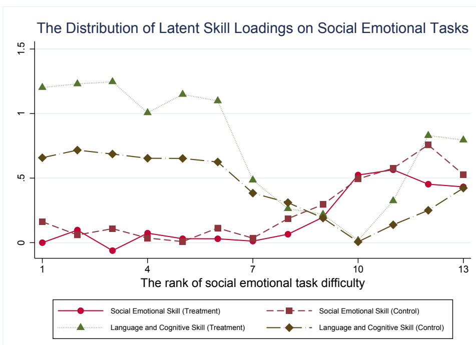
(a)



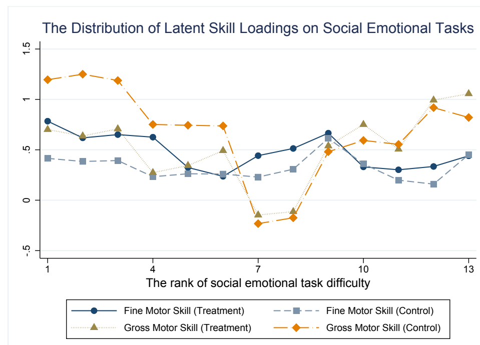
(b)



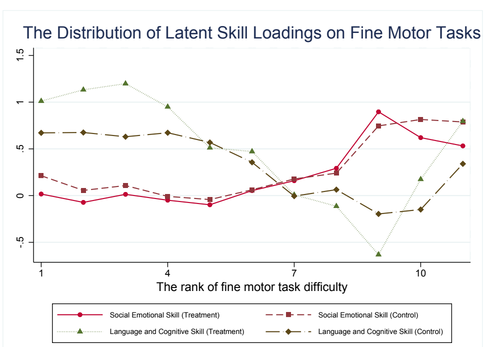
(c)



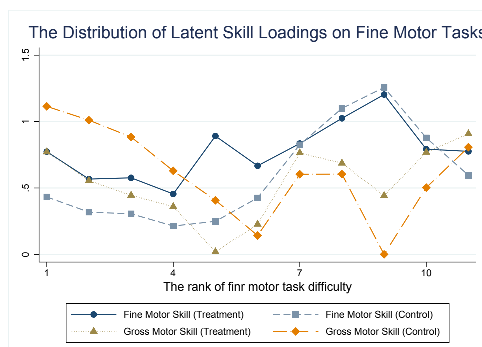
(d)



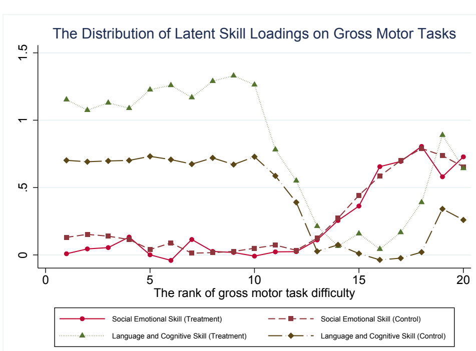
(e)



(f)



(g)



(h)

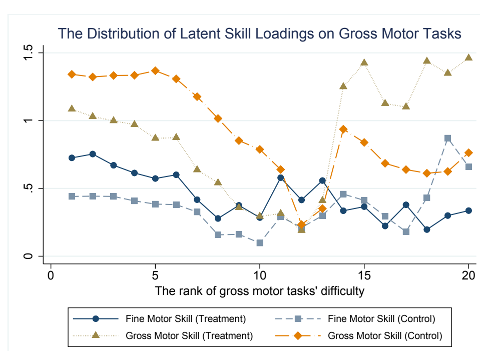
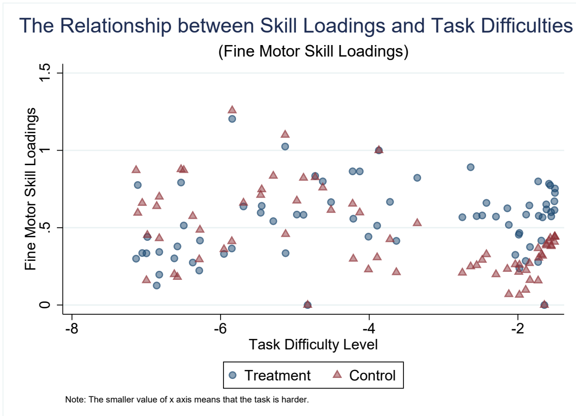
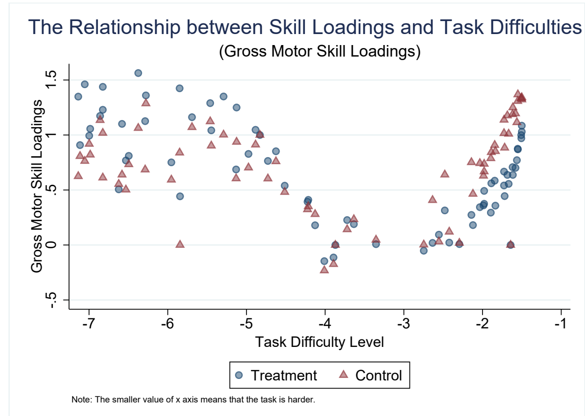


Figure Q.2: The Relationship between Latent Skill Loadings and Task Difficulties

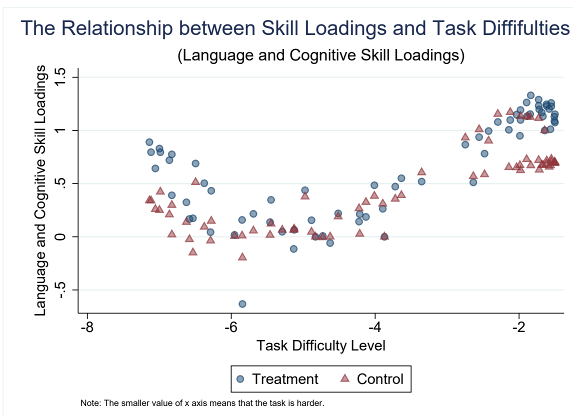
(a)



(b)



(c)



(d)

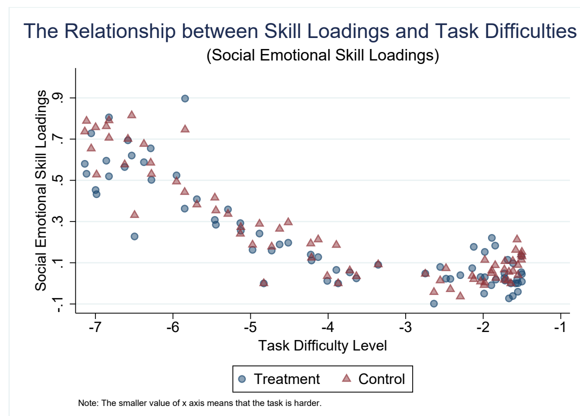


Table Q.1: The Test of Equality in the Loadings between Treatment and Control Groups ($\boldsymbol{\alpha^{j,1} = \alpha^{j,0}}$)

	Social-Emotional Skills	Fine Motor Skills	Language and Cognitive Skills	Gross Motor Skills
	Step Down p-values			
	Social-Emotional Tasks			
Play Ball with Examiner	0.147	**0.042**	**0.028**	**0.032**
Help in House	0.193	**0.049**	**0.024**	0.131
Drink From Cup	0.054	**0.038**	**0.021**	0.059
Feed Doll	0.497	**0.021**	**0.031**	**0.019**
Use Spoon/fork	0.199	0.195	**0.023**	0.390
Remove Garment	0.121	0.239	**0.026**	**0.036**
Put on Clothing	0.152	**0.018**	0.066	**0.039**
Brush Teeth with Help	**0.027**	0.087	0.153	**0.045**
Name Friend	**0.047**	0.230	0.361	0.312
Put on T-shirt	0.519	0.437	0.390	**0.031**
Dress No Help	0.365	0.290	0.201	0.362
Play Board/Card Games	0.238	0.119	0.182	0.383
Brush Teeth no Help	0.476	0.436	0.511	0.433
$\chi^2(13)$	3.686	10.156	25.010	28.250
p-value	0.994	0.681	0.023	0.008
	Language and Cognitive Tasks			
DaDa/MaMa Specific	0.336	**0.026**	0.098	**0.030**
One Word	0.162	**0.030**	0.270	0.176
Two Words	0.051	**0.025**	0.447	0.167
3 Words	0.051	**0.015**	0.130	**0.041**
6 Words	**0.033**	**0.013**	0.109	0.174
Point 2 Pictures	0.108	**0.012**	0.124	0.284
Body Parts 6	0.076	**0.034**	0.081	**0.045**
Name 1 Picture	0.431	**0.035**	0.109	0.073
Speech Half Understandable	0.061	**0.018**	0.181	0.061
Point 4 Pictures	0.454	0.112	0.510	0.091
Know 2 Actions	**0.048**	0.067	**0.046**	**0.035**
Name 4 Pictures	0.072	0.161	0.057	**0.049**
Name 1 Color	0.167	0.150	0.450	**0.037**
Use of 2 Objects	0.102	0.107	0.164	**0.043**
Speech all Understandable	0.059	0.387	0.416	0.060
Know 2 Adjectives	**0.045**	0.393	0.332	0.059
Use of 3 Objects	0.206	0.101	0.602	**0.018**
Count 1 Block	0.072	0.458	0.469	**0.041**
Know 4 Actions	0.153	0.463	0.510	0.063
Understand 4 Prepositions	0.084	0.216	0.253	0.135
Know 3 Adjectives	0.198	0.187	0.091	0.109
Opposites 2	0.318	0.211	0.060	0.303
Count 5 Blocks	0.298	0.253	0.407	**0.038**
Name 4 Colors	0.078	0.134	0.051	0.200
$\chi^2(24)$	6.252	54.315	36.339	25.651
p-value	0.999	0.000	0.050	0.371

Table Q.2: The Test of Equality in the Loadings between Treatment and Control Groups ($\boldsymbol{\alpha}^{j,1} = \boldsymbol{\alpha}^{j,0}$)

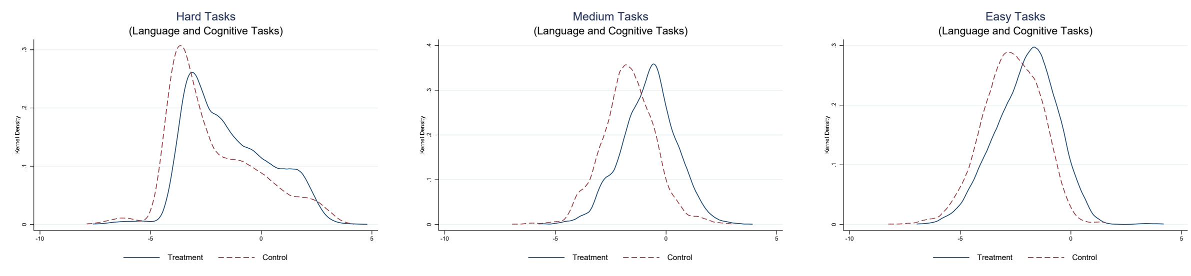
	Social-Emotional Skills	Fine Motor Skills	Language and Cognitive Skills	Gross Motor Skills
	Step Down p-values			
	Fine Motor Tasks			
Scribbles	0.162	**0.033**	**0.034**	**0.037**
Tower of 2 Cubes	0.058	0.060	**0.031**	0.429
Dump Raisin Demonstrated	0.131	**0.032**	**0.020**	0.090
Tower of 4 Cubes	0.428	**0.050**	**0.032**	0.073
Tower of 6 Cubes	0.146	**0.031**	0.248	**0.010**
Tower of 8 Cubes	0.274	**0.016**	0.219	**0.035**
Thumb Wiggle	0.173	0.613	0.668	**0.039**
Pick Longer Line	0.080	0.284	0.218	0.103
Copy Circle	**0.039**	0.371	0.481	**0.043**
Copy +	0.165	0.514	**0.026**	0.147
Draw Person 3 Parts	0.120	0.310	0.162	0.331
$\chi^2(11)$	2.774	16.364	25.894	32.449
p-value	0.993	0.128	0.007	0.001
	Gross Motor Tasks			
Get to Sitting	0.344	0.090	**0.048**	0.101
Pull to Stand	0.564	0.085	0.052	0.088
Stand 2 Seconds	0.607	0.087	**0.040**	0.060
Stand 10 Seconds	0.555	0.066	**0.036**	**0.047**
Stoop and Recover	0.180	0.052	**0.025**	0.236
Walk Well	0.093	**0.044**	**0.023**	0.051
Walk Backwards	0.075	0.087	**0.024**	0.421
Runs	0.370	**0.048**	**0.019**	0.445
Walk up Steps	0.383	**0.036**	**0.018**	0.052
Kick Ball Forward	0.128	**0.028**	**0.020**	0.254
Throw Ball Overhand	0.322	**0.019**	0.099	0.084
Jump up	0.239	**0.016**	0.095	0.087
Balance Each Foot 1 Second	0.203	**0.013**	0.239	0.396
Balance Each Foot 2 Second	0.241	0.342	0.249	0.375
Balance Each Foot 3 Second	0.045	0.158	**0.031**	0.044
Hops	0.152	0.521	0.073	0.161
Balance Each Foot 4 Second	0.674	0.133	**0.030**	**0.017**
Balance Each Foot 5 Second	0.395	0.509	0.159	**0.014**
Heel-to-toe Walk	0.688	0.358	0.076	0.347
Balance Each Foot 6 Seconds	0.477	0.101	0.059	0.351
$\chi^2(20)$	2.090	18.406	47.627	92.230
p-value	0.999	0.561	0.000	0.000
	Joint test for all tasks			
$\chi^2(68)$	14.802	99.241	134.870	178.581
p-value	0.999	0.018	0.000	0.000

Q.2 The Endline Distributions of Effective Skills $\left[(\boldsymbol{\theta}_i^d)'\boldsymbol{\alpha}^{j_k,d}\right]$

Easy tasks are defined as the bottom 33% ordered by difficulty level estimates within the same skill type. For example, there are 28 language and cognitive tasks. The top 9 language tasks ordered by difficulty parameters are defined as easy tasks. The order between 10 and 18 language tasks are defined as medium tasks. The order between 19 and 28 are defined as hard tasks. The order is within the same skill type.

Figure Q.3: Endline Distributions of Effective Skills ($\left[(\boldsymbol{\theta}_i^d)'\boldsymbol{\alpha}^{j_k,d}\right]^{\dagger}$) for Language and Cognitive Tasks

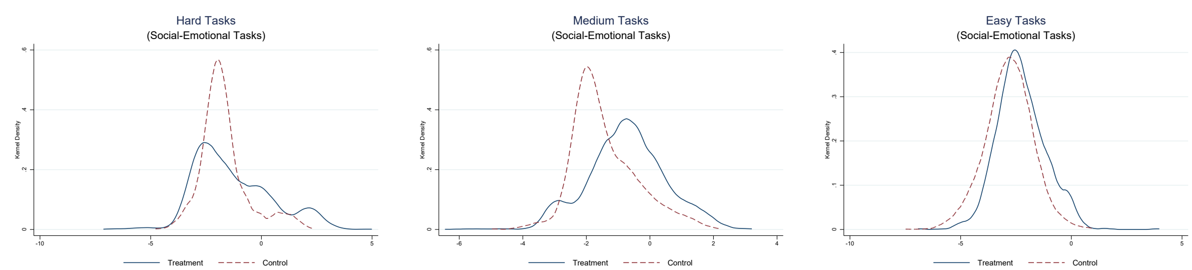
(a) Hard Tasks (b) Medium Tasks (c) Easy Tasks



† Easy tasks are defined as the bottom 33% of all language and cognitive tasks ordered by difficulty level estimates, medium tasks are those that fall between 33% and 66% of all the language and cognitive tasks ordered by difficulty level estimates, and hard tasks are the top 66% of all the language and cognitive tasks ordered by difficulty level estimates.

Figure Q.4: Endline Distributions of Effective Skills ($\left[(\boldsymbol{\theta}_i^d)'\boldsymbol{\alpha}^{j_k,d}\right]^{\dagger}$) for Social-Emotional Tasks

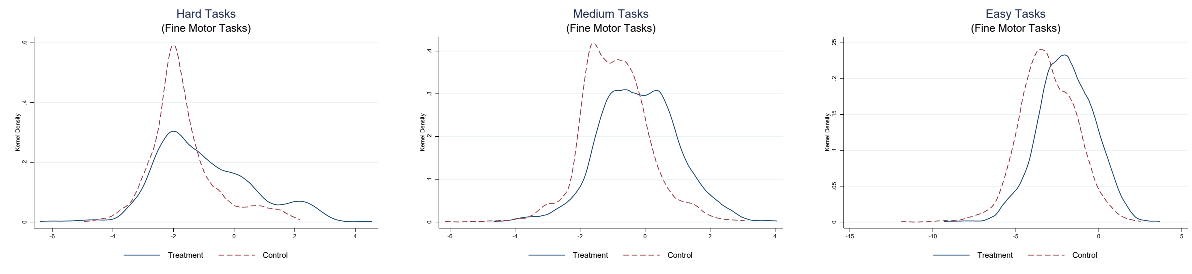
(a) Hard Tasks (b) Medium Tasks (c) Easy Tasks



† Easy tasks are defined as the bottom 33% of all social-emotional tasks ordered by difficulty level estimates, medium tasks are those that fall between 33% and 66% of all the social-emotional tasks ordered by difficulty level estimates, and hard tasks are the top 66% of all the social-emotional tasks ordered by difficulty level estimates.

Figure Q.5: Endline Distributions of Effective Skills ($\left[(\boldsymbol{\theta}_i^d)'\boldsymbol{\alpha}^{j_k,d}\right]^\dagger$) for Fine Motor Tasks

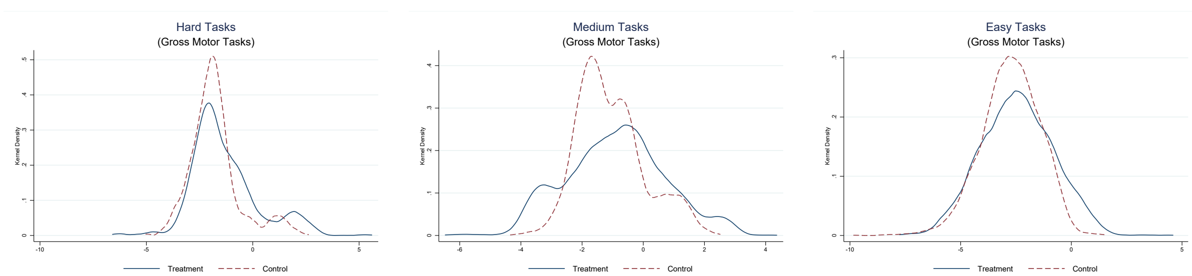
(a) Hard Tasks (b) Medium Tasks (c) Easy Tasks



† Easy tasks are defined as the bottom 33% of all fine motor tasks ordered by difficulty level estimates, medium tasks are those that fall between 33% and 66% of all the fine motor tasks ordered by difficulty level estimates, and hard tasks are the top 66% of all the fine motor tasks ordered by difficulty level estimates.

Figure Q.6: Endline Distributions of Effective Skills ($\left[(\boldsymbol{\theta}_i^d)'\boldsymbol{\alpha}^{j_k,d}\right]^\dagger$) for Gross Motor Tasks

(a) Hard Tasks (b) Medium Tasks (c) Easy Tasks



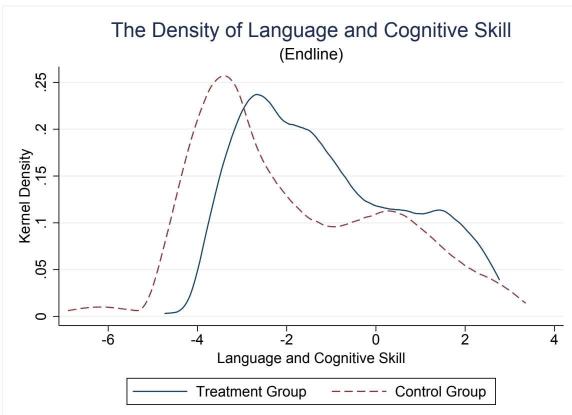
† Easy tasks are defined as the bottom 33% of all gross motor tasks ordered by difficulty level estimates, medium tasks are those that fall between 33% and 66% of all the gross motor tasks ordered by difficulty level estimates, and hard tasks are the top 66% of all the gross motor tasks ordered by difficulty level estimates.

R Densities and Cumulative Density Functions of Estimated Skill Distributions

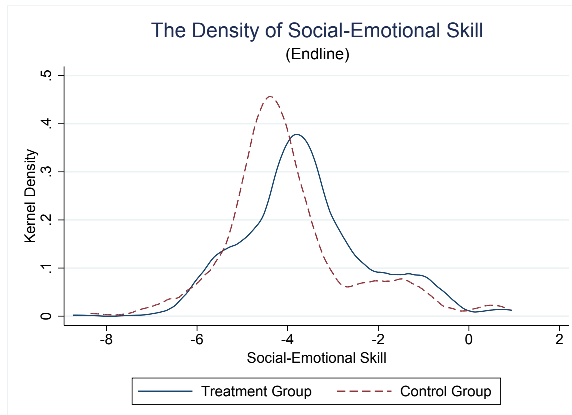
We plot both densities and cdfs of the estimated latent skills.

Figure R.1: Treated and Untreated Skill Distribution
(Endline)

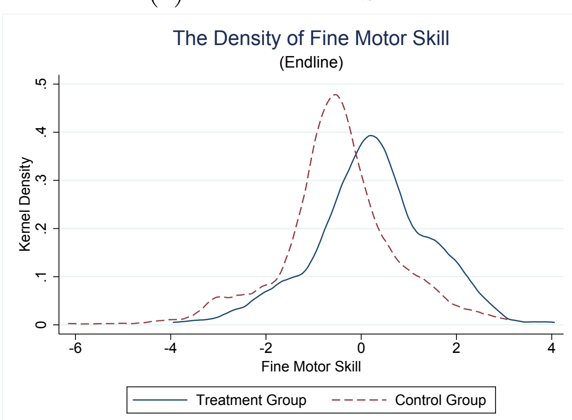
(a) Language and Cognitive Skills



(b) Social-Emotional Skills



(c) Fine Motor Skills



(d) Gross Motor Skills

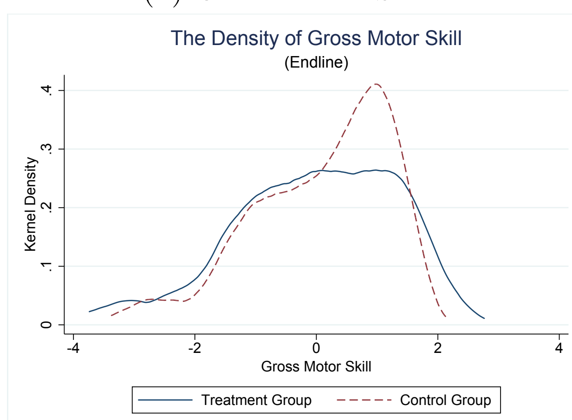
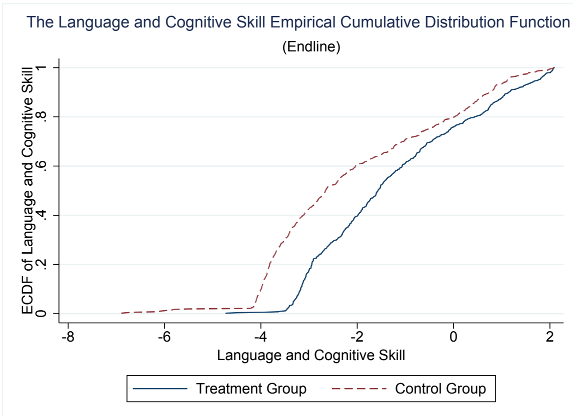
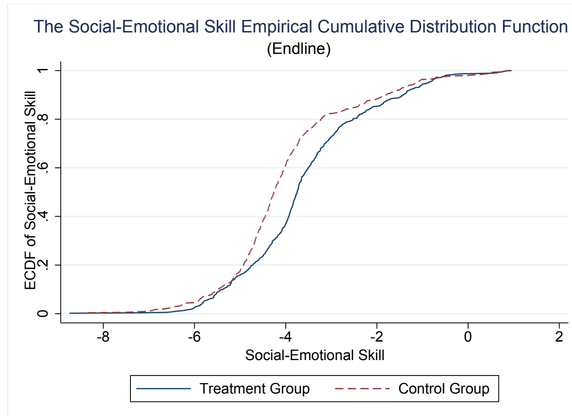


Figure R.2: Treated and Untreated Cumulative Density Functions (Endline)

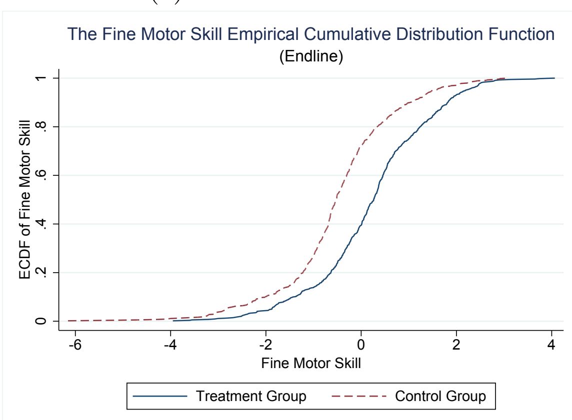
(a) Language and Cognitive Skills



(b) Social-Emotional Skills



(c) Fine Motor Skills



(d) Gross Motor Skills

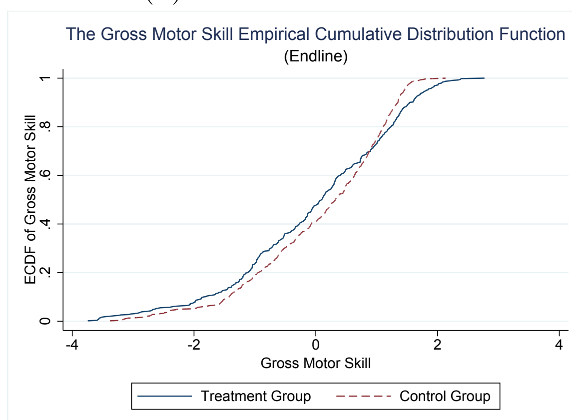
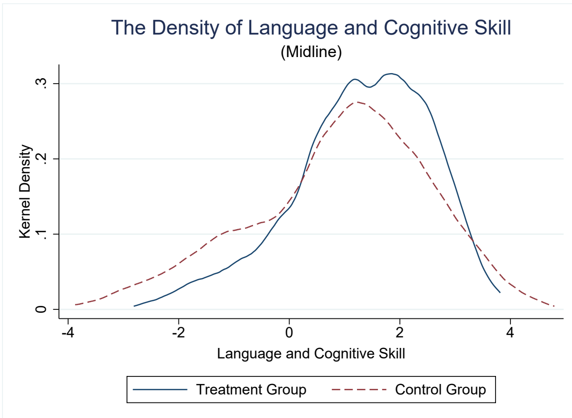
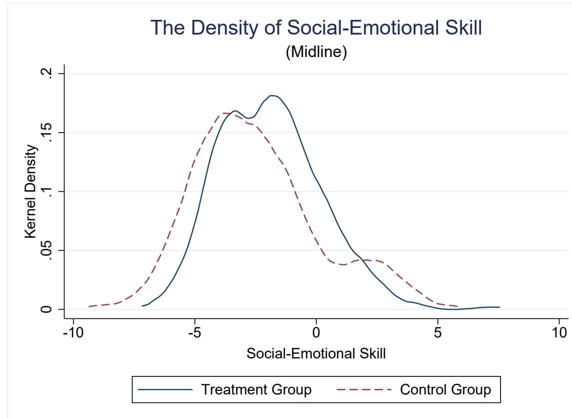


Figure R.3: Treated and Untreated Skill Distribution
(Midline)

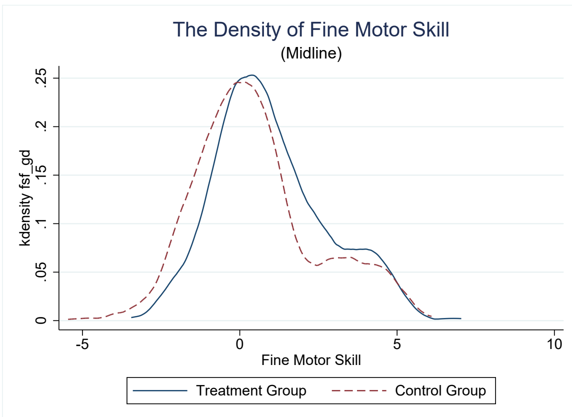
(a) Language and Cognitive Skills



(b) Social-Emotional Skills



(c) Fine Motor Skills



(d) Gross Motor Skills

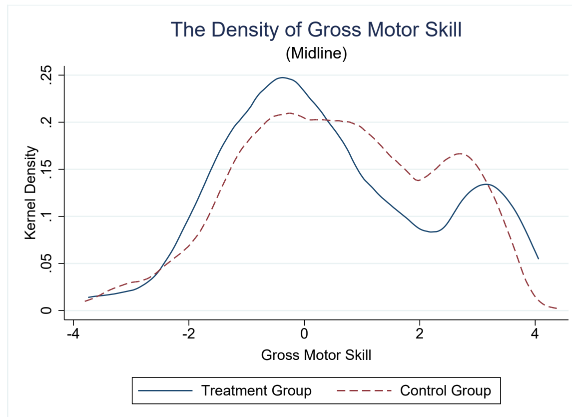
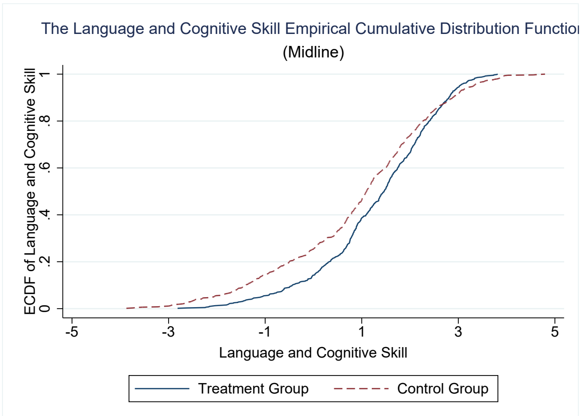
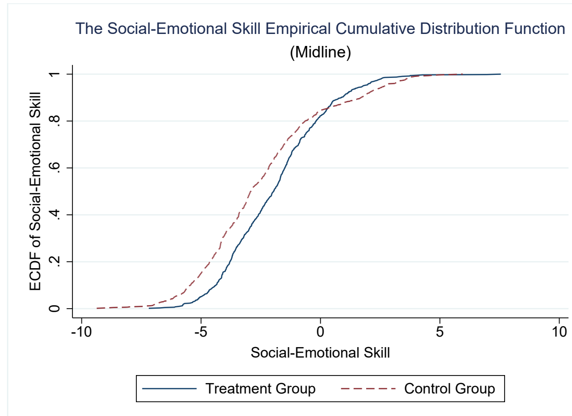


Figure R.4: Treated and Untreated Cumulative Density Functions
(Midline)

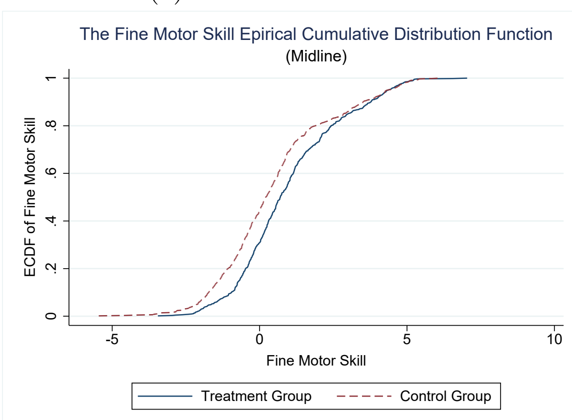
(a) Language and Cognitive Skills



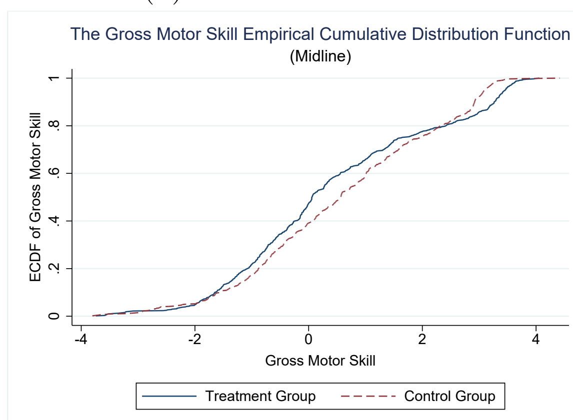
(b) Social-Emotional Skills



(c) Fine Motor Skills



(d) Gross Motor Skills



S Stochastic Dominance

We test for stochastic dominance of the estimated skill curves for treatments and controls. Figure S.1, the generalized Lorenz curve, shows the average cumulative values at each cumulative proportion observation. At each cumulative proportion, the treated children have higher language skills. Similarly, Figure S.1 gives the maximum language and cognitive skill values at each percentile. It is clear that the treated group has larger language skill values at each percentile. Figures S.2–S.4 show the same measures for social-emotional, fine motor and gross motor skills, respectively. We can find similar patterns for social-emotional and fine motor skills but not for gross motor skill.

Figure S.1: Language and Cognitive Skills Stochastic Dominance Curves

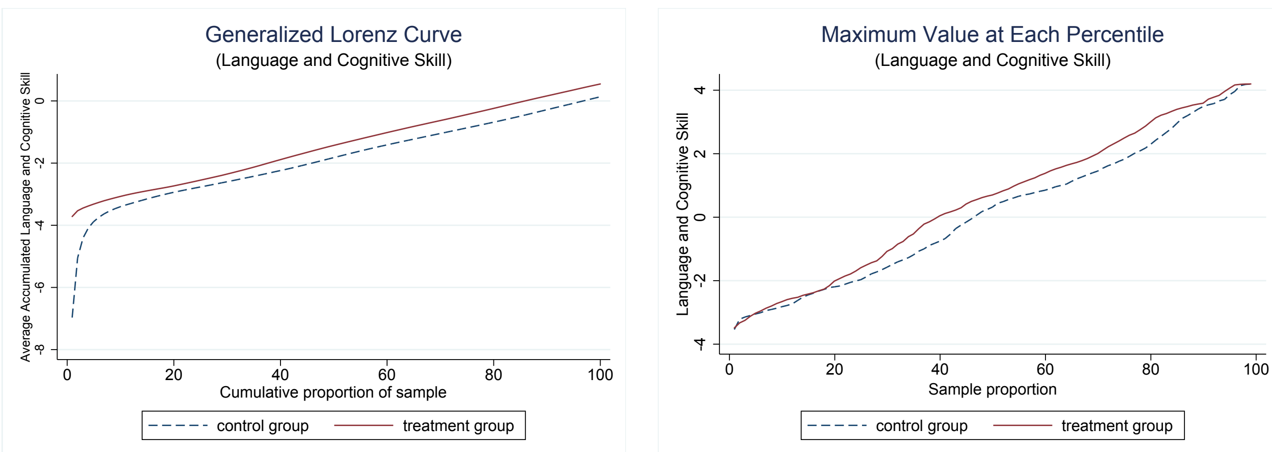


Figure S.2: Social-Emotional Skills Stochastic Dominance Curves

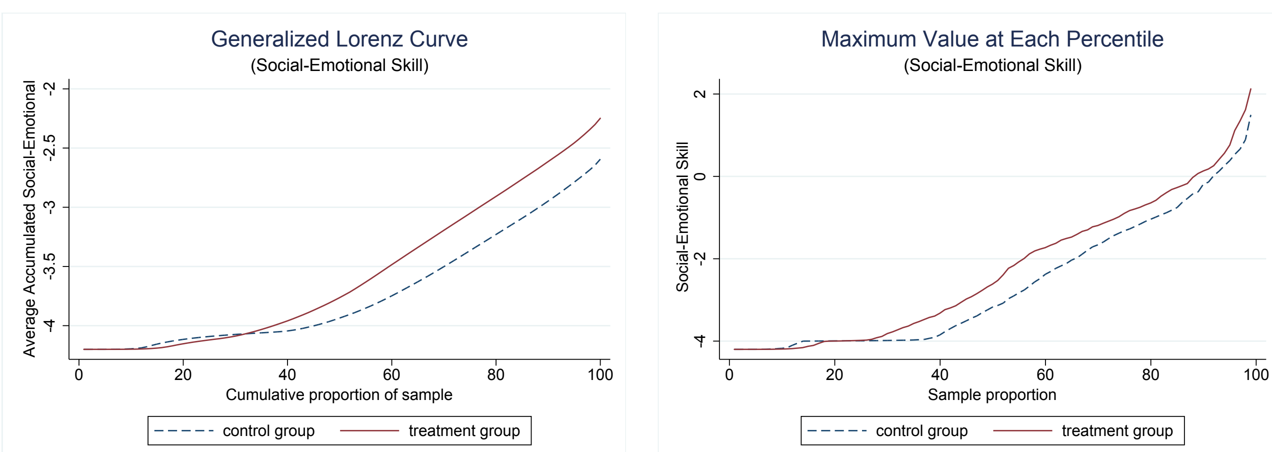
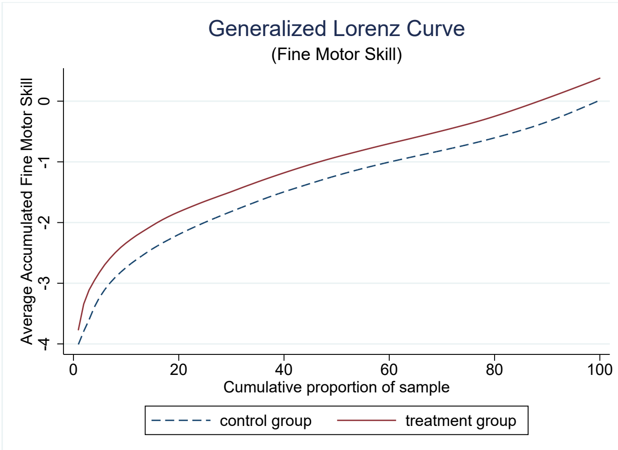


Figure S.3: Fine Motor Skills Stochastic Dominance Curves



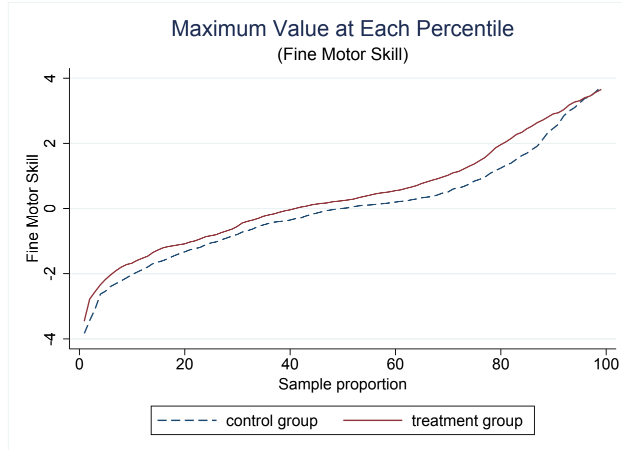
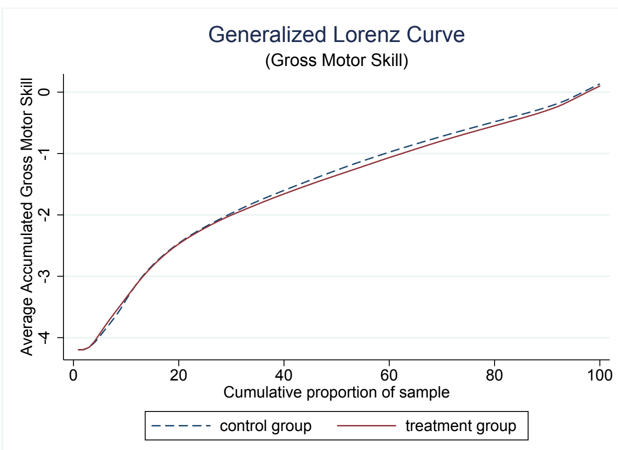
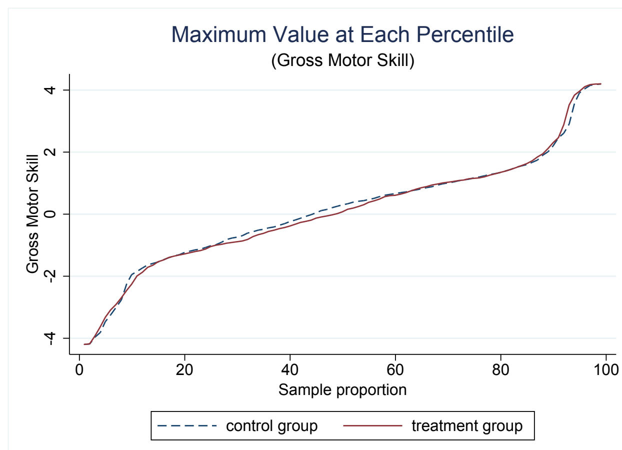


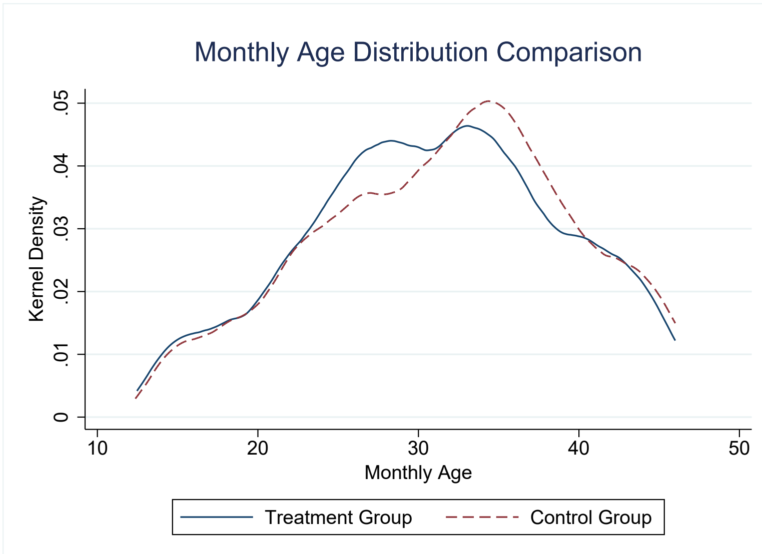
Figure S.4: Gross Motor Skills Stochastic Dominance Curves





T Monthly Age Distribution Comparison

Figure T.1: Monthly Age Distribution Comparison



Note: The p-value of the Kruskal-Wallis test is 0.18.

U Changing the Order of the Decomposition

U.1 Decompose Skill Loadings First

We try different orders of decomposition. Table U.1 shows quantitatively similar decomposition results regardless of the order used.

Table U.1: Source of Treatment Effects (Decompose Skill Loadings First)

	Total Net Treatment Effects	From Observable Covariates	From Skill Loadings	From Latent Skills
Language and Cognitive	1.143	0.100	0.060	0.984
	(0.185)	(0.190)	(0.193)	(0.188)
		9%	5%	86%
Social-Emotional	0.239	-0.196	0.082	0.354
	(0.083)	(0.084)	(0.085)	(0.084)
		-82%	34%	148%
Fine Motor	0.317	-0.096	0.077	0.336
	(0.085)	(0.087)	(0.089)	(0.088)
		-30%	24%	106%
Gross Motor	0.164	-0.103	0.111	0.156
	(0.100)	(0.106)	(0.105)	(0.103)
		-63%	68%	95%

Notes: 1. Total treatment effects for skill k are $\frac{1}{N_{J_k}} \sum_{j_k=1}^{N_{J_k}} \left(\frac{\sum_{i=1}^{N_I} Y^{j_k,i} D_i}{\sum_{i=1}^{N_I} D_i} - \frac{\sum_{i=1}^{N_I} Y^{j_k,i}(1-D_i)}{\sum_{i=1}^{N_I}(1-D_i)} \right)$ assuming both denominators are nonzero and N_I is the number of observations.
2. To ensure that the observed covariates are balanced between the treatment and control groups, we consider the sample of children who are younger than 46 months and older than 12 months.
3. Standard errors are reported in parentheses.

U.2 Decomposition Results based on Estimates Controlling for Family Background Covariates

In this section, we report the decomposition results based on the estimates conditional on more family background covariates. Our findings are similar: the order of decomposition does not matter much.

Table U.2: Source of Treatment Effects (Decompose Observed Covariates First)
(Estimates Controlling for Family Background Covariates)

Tasks	Total Net Treatment Effects	From Skill Loadings	From Observable Covariates	From Latent Skills
Language and Cognitive	1.152	-0.988	0.462	1.678
	(0.234)	(0.237)	(0.224)	(0.247)
		-86%	40%	146%
Social-Emotional	0.158	-0.504	0.145	0.517
	(0.081)	(0.086)	(0.076)	(0.091)
		-320%	92%	328%
Fine Motor	0.305	-0.502	0.196	0.611
	(0.104)	(0.101)	(0.094)	(0.109)
		-164%	64%	200%
Gross Motor	0.173	-0.675	0.217	0.631
	(0.129)	(0.134)	(0.120)	(0.141)
		-391%	126%	365%

Notes: 1. Total treatment effects for skill k are $\frac{1}{N_{J_k}}\sum_{j_k=1}^{N_{J_k}}\left(\frac{\sum_{i=1}^{N_I} Y^{j_k,i}D_i}{\sum_{i=1}^{N_I} D_i} - \frac{\sum_{i=1}^{N_I} Y^{j_k,i}(1-D_i)}{\sum_{i=1}^{N_I}(1-D_i)}\right)$ assuming both denominators are nonzero and N_I is the number of observations.
2. To ensure that the observed covariates are balanced between the treatment and control groups, we consider the sample of children who are younger than 46 months and older than 12 months.
3. Standard errors are reported in parentheses.

Table U.3: Source of Treatment Effects (Decompose Skill Loadings First)
(Estimates Controlling for Family Background Covariates)

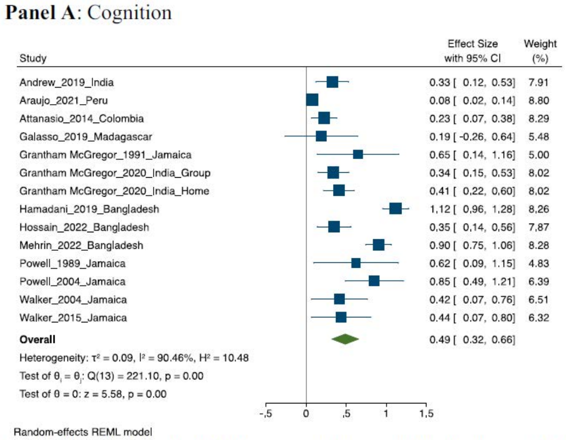
Tasks	Total Net Treatment Effects	From Skill Loadings	From Observable Covariates	From Latent Skills
Language and Cognitive	1.152	-1.069	0.543	1.678
	(0.234)	(0.222)	(0.235)	(0.247)
		-93%	47%	146%
Social-Emotional	0.158	-0.594	0.234	0.517
	(0.081)	(0.077)	(0.087)	(0.091)
		-377%	149%	328%
Fine Motor	0.305	-0.516	0.211	0.611
	(0.104)	(0.093)	(0.100)	(0.109)
		-169%	69%	200%
Gross Motor	0.173	-0.738	0.280	0.631
	(0.129)	(0.128)	(0.141)	(0.141)
		-427%	162%	365%

Notes: 1. Total treatment effects for skill k are $\frac{1}{N_{J_k}}\sum_{j_k=1}^{N_{J_k}}\left(\frac{\sum_{i=1}^{N_I} Y^{j_k,i}D_i}{\sum_{i=1}^{N_I} D_i} - \frac{\sum_{i=1}^{N_I} Y^{j_k,i}(1-D_i)}{\sum_{i=1}^{N_I}(1-D_i)}\right)$ assuming both denominators are nonzero and N_I is the number of observations.
2. To ensure that the observed covariates are balanced between the treatment and control groups, we consider the sample of children who are younger than 46 months and older than 12 months.
3. Standard errors are reported in parentheses.

V Previous Studies of the Jamaica Program

This appendix reports empirical results for endline studies that attempt to replicate the Jamaica curriculum. We report estimates documented in Jervis et al. (2023), an extensive meta-analysis. These estimates include the unadjusted estimates reported in this paper. We report endline unadjusted (for difficulty) treatment effects which are standard in the literature. None of these papers adjusts for item difficulty or estimates distributions of latent skills as we do in this paper. We note that adjustment for difficulty vitally affects treatment effects. Our unadjusted treatment effects are similar to those in the literature. For all except home stimulation and motor skills, we note the usual reservations about meta-analysis. The socioeconomic characteristics of the populations in the studies are not necessarily comparable, the survey instruments used are not necessarily comparable and the follow up periods are not necessarily comparable. In addition, the versions of the Jamaican program used are not in general identical and implementation is done in various modes: (a) pure home visits; (b) center visits; and (c) hybrids.

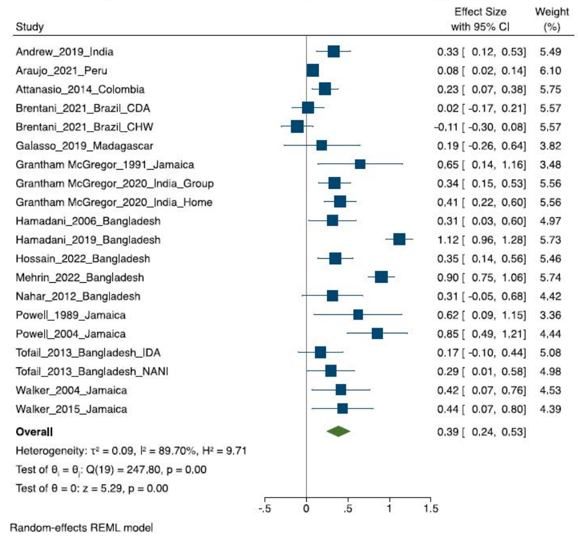
Panel A: Cognition



Note: Total number of trials is 14. Bayley Scales of Infant and Toddler Development (Bayley)-III is present in seven trials, Griffiths Scales of Mental Development (Griffiths) in five and Ages and Stages Questionnaire (ASQ) in two trials. Two trials are internally standardized.

Source: Jervis et al. (2023)

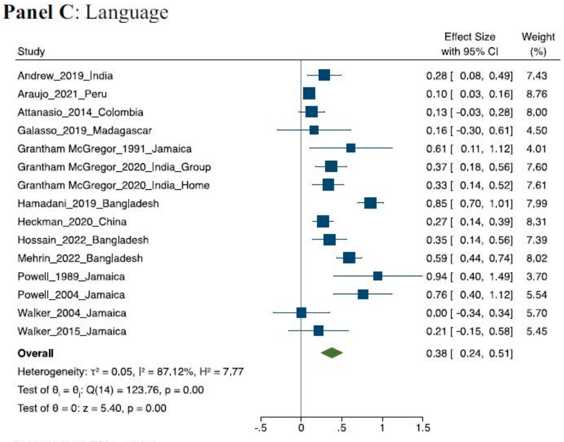
Panel B: Cognition and combined cognition and language



Note: Total number of trials is 20. Bayley-III is present in seven trials, Bayley-II in four, Griffiths in five, ASQ in two and Regional Project on Child Development Indicators (PRIDI) in two trials. Four trials are internally standardized.

Source: Jervis et al. (2023)

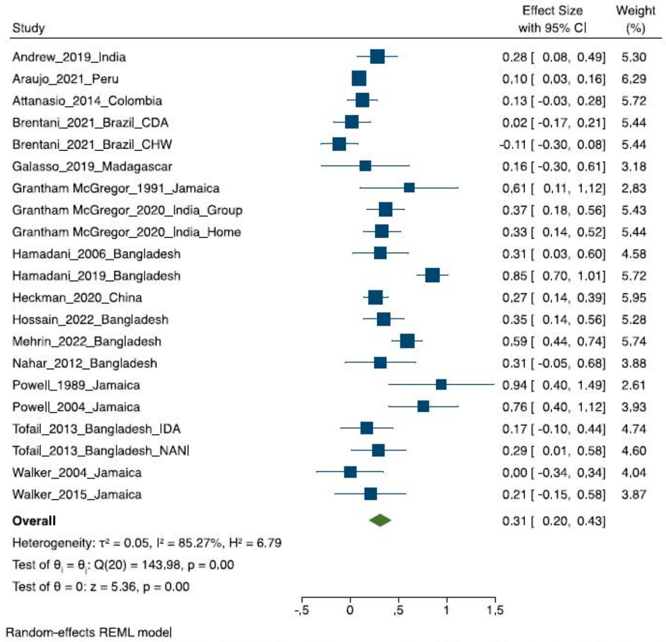
Panel C: Language



Note: Total number of trials is 15. Bayley-III is present in seven trials, Griffiths in five, ASQ in two and Denver Developmental Screening Test (Denver)-II in one trial. Three trials are internally standardized.

Source: Jervis et al. (2023)

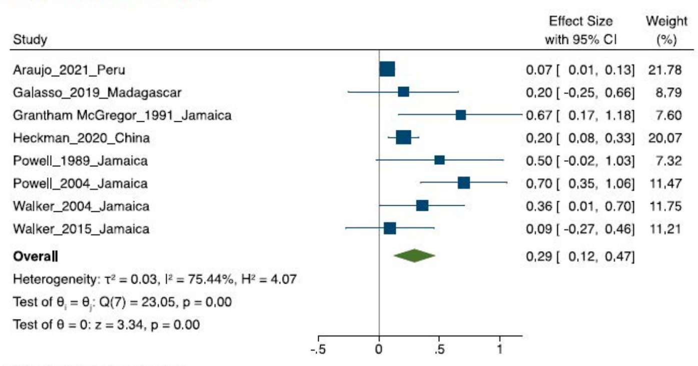
Panel D: Language and combined cognition and language



Note: Total number of trials is 21. Bayley-III is present in trials studies, Bayley-II in four, Griffiths in five, ASQ in two, PRIDI in two and Denver-II in one trial. Five trials are internally standardized.

Source: Jervis et al. (2023)

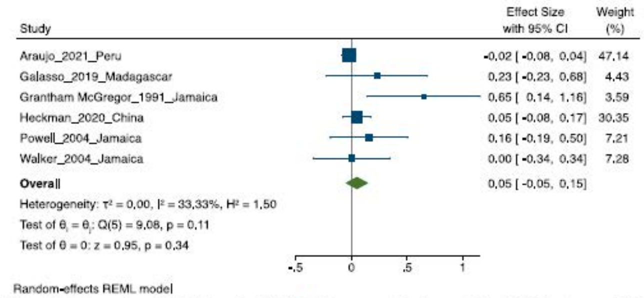
Panel E: Fine Motor



Note: Total number of trials is eight. Griffiths is present in five trials, ASQ in two and Denver-II in one trial. Three trials are internally standardized.

Source: Jervis et al. (2023)

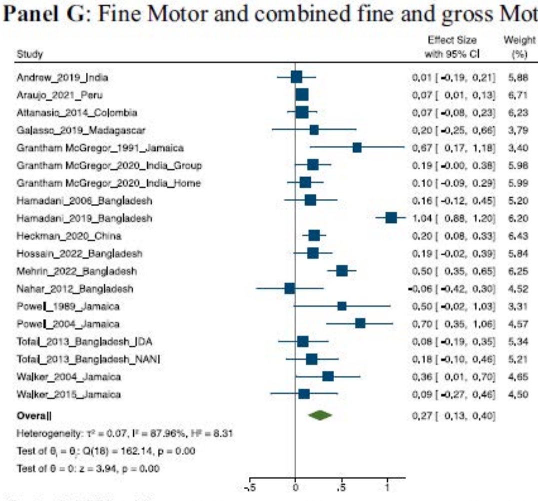
Panel F: Gross Motor



Note: Total number of trials is six. Griffiths is present in three trials, ASQ in two and Denver-II in one trial. Three trials are internally standardized.

Source: Jervis et al. (2023)

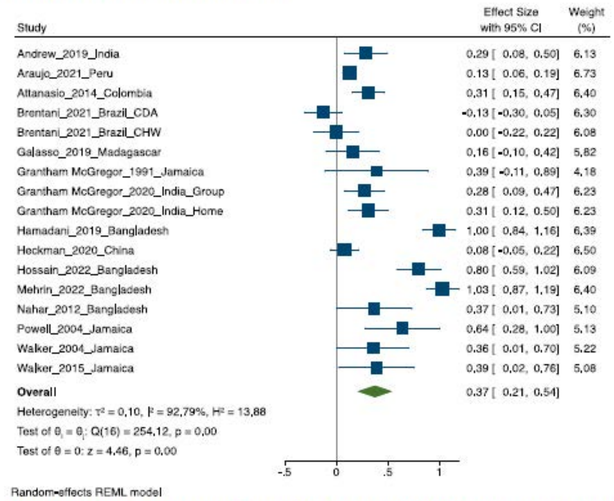
Panel G: Fine Motor and combined fine and gross Motor



Note: Total number of trials is 19. Bayley-III is present in seven trials, Bayley-II in four, Griffiths in five, ASQ in two and Denver-II in one trial. Three trials are internally standardized.

Source: Jervis et al. (2023)

Panel H: Home Stimulation



Note: Total number of trials is 17. Home Observation for Measurement of the Environment (HOME) is present in ten trials and UNICEF's Family Care Indicators (FCI) in seven trials.

Source: Jervis et al. (2023)

References

Abadie, A. and G. W. Imbens (2012). A martingale representation for matching estimators. *Journal of the American Statistical Association 107*(498), 833–843.

Aboud, F. and A. Yousafzai (2015, 1). Global health and development in early childhood. *Annual Review of Psychology 66*(1), 433–457.

Agostinelli, F. and M. Wiswall (2024). Estimating the technology of children's skill formation. *Journal of Political Economy* (No. w22442).

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume 5, Berkeley, CA, pp. 111–150. University of California Press.

Andrew, A., O. Attanasio, B. Augsburg, M. Day, S. Grantham-mcgregor, C. Meghir, F. Mehrin, S. Pahwa, and M. Rubio-codina (2020). Effects of a scalable home-visiting intervention on child development in slums of urban india: evidence from a randomised controlled trial. *Journal of Child Psychology and Psychiatry 61*(6), 644–652.

Andrew, A., O. Attanasio, E. Fitzsimons, S. Grantham-Mcgregor, C. Meghir, and M. Rubio-Codina (2018, 4). Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in colombia. *PLOS Medicine 15*(4), e1002556.

Araujo, M., M. Dormal, S. Grantham-Mcgregor, F. Lazarte, M. Rubio-Codina, and N. Schady (2021). Home visiting at scale and child development. *Journal of Public Economics Plus 2*, 100003.

Attanasio, O., H. Baker-Henningham, R. Bernal, C. Meghir, D. Pineda, and M. Rubio-Codina (2022). Early stimulation and nutrition: The impacts of a scalable intervention. *J Eur Econ Assoc 20*(4), 1395–1432.

Attanasio, O., K. Meghir, M. Rubio-Codina, K. Welch, and S. Quddus (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in colombia: cluster randomized controlled trial. *BMJ 349*, 5785.

Brentani, A., S. Walker, S. Chang-Lopez, S. Grisi, C. Powell, and G. Fink (2021, 1). A home visit-based early childhood stimulation programme in brazil—a randomized controlled trial. *Health Policy and Planning 36*(3), 288–297.

Britto, P., N. Ulkuer, W. Hodges, and M. Mccarthy (2017). *Global Policy Landscape and Early Childhood Development*, Volume 389, pp. 65–82. Oxford University PressNew York.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Chamberlain, G. (1977a, March). Education, income, and ability revisited. *Journal of Econometrics 5*(2), 241–257.

Chamberlain, G. (1977b). An instrumental variable interpretation of identification in variance components and mimic models. In P. Taubman (Ed.), *Kinometrics: Determinants of Socioeconomic Success Within and Between Families*. New York: North-Holland.

Chamberlain, G. and Z. Griliches (1975, June). Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers. *International Economic Review 16*(2), 422–449.

Chen, M., I. Fernández-Val, and M. Weidner (2021). Nonlinear factor models for network and panel data. *Journal of Econometrics 220*(2), 296–324.

Emmers, D., Q. Jiang, and H. Xue (2021). Early childhood development and parental training interventions in rural china: a systematic review and meta-analysis. *BMJ Glob Health 6*(8), 5578.

Galasso, E., A. Weber, C. Stewart, L. Ratsifandrihamanana, and L. Fernald (2019, 9). Effects of nutritional supplementation and home visiting on growth and development in young children in madagascar: a cluster-randomised controlled trial. *The Lancet Global Health 7*(9), e1257–e1268.

Goldberger, A. S. (1972, November). Structural equation methods in the social sciences. *Econometrica 40*(6), 979–1001.

Grantham-Mcgregor, S., A. Adya, O. Attanasio, B. Augsburg, J. Behrman, B. Caeyers, M. Day, P. Jervis, R. Kochar, P. Makkar, C. Meghir, A. Phimister, M. Rubio-Codina, and K. Vats (2020, 12). Group sessions or home visits for early childhood development in india: A cluster rct. *Pediatrics 146*(6).

Grantham-Mcgregor, S., C. Powell, S. Walker, and J. Himes (1991, 7). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the jamaican study. *The Lancet 338*(8758), 1–5.

Grantham-Mcgregor, S. and J. Smith (2016). Extending the jamaican early childhood development intervention. *Journal of Applied Research on Children: Informing Policy for Children at Risk 7*(2), 4.

Hamadani, J., S. Huda, F. Khatun, and S. Grantham-Mcgregor (2006, 10). Psychosocial stimulation improves the development of undernourished children in rural bangladesh. *The Journal of Nutrition 136*(10), 2645–2652.

Hamadani, J., S. Mehrin, F. Tofail, M. Hasan, S. Huda, H. Baker-Henningham, D. Ridout, and S. Grantham-Mcgregor (2019, 3). Integrating an early childhood development programme into bangladeshi primary health-care services: an open-label, cluster-randomised controlled trial. *The Lancet Global Health 7*(3), e366–e375.

Heckman, J., B. Liu, M. Lu, and J. Zhou (2023). The impacts of a prototypical home visiting program on child skills. *National Bureau of Economic Research*.

Higgins, J., J. Thomas, and J. Chandler (2019). Cochrane handbook for systematic reviews of interventions, 2nd ed.

Hossain, S., B. Roy, H. Sujon, T. Tran, J. Fisher, F. Tofail, S. El Arifeen, and J. Hamadani (2022, 1). Effects of integrated psychosocial stimulation (ps) and unconditional cash transfer (uct) on children's development in rural bangladesh: A cluster randomized controlled trial. *Social Science and Medicine 293*.

Jeong, J., E. Franchett, C. Ramos De Oliveira, K. Rehmani, and A. Yousafzai (2021, 5). Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLOS Medicine 18*(5), e1003602.

Jeong, J., J. Obradović, M. Rasheed, D. Mccoy, G. Fink, and A. Yousafzai (2019, 1). Maternal and paternal stimulation: Mediators of parenting intervention effects on preschoolers' development. *Journal of Applied Developmental Psychology 60*, 105–118.

Jeong, J., H. Pitchik, and G. Fink (2021, 3). Short-term, medium-term and long-term effects of early parenting interventions in low- and middle-income countries: a systematic review. *BMJ Global Health 6*(3), e004067.

Jeong, J., H. Pitchik, and A. Yousafzai (2018, 4). Stimulation interventions and parenting in low- and middle-income countries: A meta-analysis. *Pediatrics 141*(4).

Jervis, P., J. Coore-Hall, H. O. Pitchik, C. D. Arnold, S. Grantham-McGregor, M. Rubio-Codina, H. Baker-Henningham, L. C. Fernald, J. Hamadani, J. A. Smith, and Others

(2023). The Reach Up parenting program, child development, and maternal depression: A meta-analysis. *Pediatrics 151*(Supplement 2).

Jöreskog, K. G. and A. S. Goldberger (1975, September). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association 70*(351), 631–639.

Luoto, J., I. Lopez Garcia, F. Aboud, D. Singla, L. Fernald, H. Pitchik, U. Saya, R. Otieno, and E. Alu (2021, 3). Group-based parenting interventions to promote child development in rural kenya: a multi-arm, cluster-randomised community effectiveness trial. *The Lancet Global Health 9*(3), e309–e319.

Maasoumi, E. and L. Wang (2019). The gender gap between earnings distributions. *Journal of Political Economy 127*(5), 2438–2504.

Maureen, M. (2017). Early childhood development coming of age: science through the life course black. *The Lancet 389*, 77–90.

Mehrin, S., M. Hasan, F. Tofail, S. Shiraji, D. Ridout, S. Grantham-Mcgregor, J. Hamadani, and H. Baker-Henningham (2022, 2). Integrating a group-based, early childhood parenting intervention into primary health care services in rural bangladesh: A cluster-randomised controlled trial. *Front Pediatr 10*, 886542.

Murray, L., L. De Pascalis, M. Tomlinson, Z. Vally, H. Dadomo, B. Maclachlan, C. Woodward, and P. Cooper (2016, 7). Randomized controlled trial of a book-sharing intervention in a deprived south african community: effects on carer–infant interactions, and their relation to infant cognitive and socioemotional outcome. *Journal of Child Psychology and Psychiatry 57*(12), 1370–1379.

Nahar, B., M. Hossain, J. Hamadani, T. Ahmed, S. Huda, S. Grantham-Mcgregor, and L. Persson (2012, 2). Effects of a community-based approach of food and psychosocial stimulation on growth and development of severely malnourished children in bangladesh: a randomised trial. *European Journal of Clinical Nutrition 66*(6), 701–709.

Palmer, F. H. (1971). *Concept training curriculum for children ages two to five*. Stony Brook, NY: State University of New York at Stony Brook.

Powell, C., H. Baker-Henningham, S. Walker, J. Gernay, and S. Grantham-Mcgregor (2004, 6). Feasibility of integrating early stimulation into primary care for undernourished jamaican children: cluster randomised controlled trial. *BMJ 329*(7457), 89.

Powell, C. and S. Grantham-Mcgregor (1989, 7). Home visiting of varying frequency and child development. *Pediatrics 84*(1), 157–164.

Schünemann, H., J. Brożek, G. Guyatt, and A. Oxman (2023, 2). Grade: assessing the quality of evidence for diagnostic recommendations. *Annals of Internal Medicine 149*(12), JC6.

Singla, D., E. Kumbakumba, and F. Aboud (2015, 8). Effects of a parenting intervention to address maternal psychological wellbeing and child development and growth in rural uganda: a community-based, cluster-randomised trial. *The Lancet Global Health 3*(8), e458–e469.

Tofail, F., J. Hamadani, F. Mehrin, D. Ridout, S. Huda, and S. Grantham-Mcgregor (2013, 6). Psychosocial stimulation benefits development in nonanemic children but not in anemic, iron-deficient children. *The Journal of Nutrition 143*(6), 885–893.

Tsiatis, A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.

Uzgiris, I. C. and J. M. Hunt (1975). *Assessment in Infancy: Ordinal Scales of Psychological Development*. University of Illinois Press.

Walker, S., S. Chang, C. Powell, and S. Grantham-Mcgregor (2004, 6). Psychosocial intervention improves the development of term low-birth-weight infants. *The Journal of Nutrition 134*(6), 1417–1423.

Walker, S., S. Chang, and J. Smith (2018). The reach up early childhood parenting program: origins, content and implementation. taking a global view on infants, toddlers and their families. *Zero Three J 38*, 37–43.

Walker, S., S. Chang, A. Wright, R. Pinto, J. Heckman, and S. Grantham-mcgregor (2022). Cognitive, psychosocial, and behaviour gains at age 31 years from the jamaica early childhood stimulation trial. *Journal of Child Psychology and Psychiatry 63*(6), 626–635.

Walker, S., S. Chang, N. Younger, and S. Grantham-mcgregor (2010, 6). The effect of psychosocial stimulation on cognition and behaviour at 6 years in a cohort of term, low-birthweight jamaican children. *Developmental Medicine and Child Neurology 52*(7), 148–e154.

Walker, S., C. Powell, S. Chang, H. Baker-Henningham, S. Grantham-Mcgregor, M. Vera-Hernández, and F. López Bóo (2023, 2). Delivering parenting interventions through health services in the caribbean.

Walker, S., T. Wachs, S. Grantham-Mcgregor, M. Black, C. Nelson, S. Huffman, H. Baker-Henningham, S. Chang, J. Hamadani, B. Lozoff, J. Gardner, C. Powell, A. Rahman, and L. Richter (2011, 10). Inequality in early childhood: risk and protective factors for early child development. *The Lancet 378*(9799), 1325–1338.

Wang, F. (2020). Maximum likelihood estimation and inference for high dimensional generalized factor models with application to factor-augmented regressions. *Journal of Econometrics*.

Williams, B. (2020). Identification of the linear factor model. *Econometric Reviews 39*(1), 92–109.

Yousafzai, A., F. Aboud, M. Nores, and R. Kaur (2018, 5). Reporting guidelines for implementation research on nurturing care interventions designed to promote early childhood development. *Annals of the New York Academy of Sciences 1419*(1), 26–37.

Yousafzai, A., J. Obradović, M. Rasheed, A. Rizvi, X. Portilla, N. Tirado-Strayer, S. Siyal, and U. Memon (2016, 8). Effects of responsive stimulation and nutrition interventions on children's development and growth at age 4 years in a disadvantaged population in pakistan: a longitudinal follow-up of a cluster-randomised factorial effectiveness trial. *The Lancet Global Health 4*(8), e548–e558.