# Interactions as Investments: The Microdynamics and Measurement of Early Childhood Learning*

James Heckman and Jin Zhou

Center for the Economics of Human Development
and Department of Economics
University of Chicago
October 31, 2021

*This research was supported in part by NIH grant NICHD R37HD065072, the Institute for New Economic Thinking and an anonymous donor. We thank our partners on the China REACH project: China Development Research Foundation, Beijing Normal University, Child Development Research Group at the University of West Indies, and Dr. Sally Grantham-McGregor. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders, partners, or the official views of the National Institutes of Health. This paper was presented as a keynote at the Africa Meeting of the Econometric Society (June 4, 2021), the China Meeting of the Econometric Society (July 1, 2021), and the Bonn IZA-Briq Life-Cycle Workshop (August 19, 2021). We have benefited from the comments of our discussant Lance Lochner at a September 2019 Becker Friedman Institute conference in Chicago, as well as those of Alison Baulos and Rebecca Harles. Erlfang Tsai, Fuyao Wang, and Alejandra Campos contributed highly competent and insightful research assistance and commentary. An online appendix can be found at https://cehd.uchicago.edu/interactions-as-investments.

**Abstract**

This paper uses novel experimental data from a widely-emulated home visiting program implemented in China with high-frequency measurements to investigate the effectiveness of different interactions in promoting child development. Quality interactions between home visitors and caregivers enhance child skill development. We propose and implement tests for dynamic complementarity that do not rely on arbitrary measures of skill. We formulate and estimate a dynamic learning model for multiple skills and quantify the sources of learning in the early lifecycle. Using our model, we test and reject the widely-held assumption of age invariance for language and cognitive skills.

**JEL Codes:** I3, J1, C5, D2, O12, C9
**Keywords:** child development, measures of skills, scaffolding, targeting, experiment

James J. Heckman
Center for the Economics
of Human Development
and Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Email: jjh@uchicago.edu

Jin Zhou
Center for the Economics
of Human Development
and Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Email: jinzhou@uchicago.edu

# 1 Introduction

The study of early childhood investment and its consequences is an active field. Many developing countries consider early childhood investment to be a valuable strategy for promoting national skill development (e.g., Engle et al., 2011; Britto et al., 2013). The search is on for effective, low-cost strategies that are adaptable to less developed settings.

This paper analyzes a low-cost home visiting program in China, based on a widely-emulated format originally developed in Jamaica that has been shown to be effective in developing child skills (e.g., Grantham-McGregor and Smith, 2016; Gertler et al., 2014, 2021). The study of home visiting effectiveness isolates a component of successful omnibus programs that include this feature (see Elango et al., 2016). We investigate mechanisms producing growth of knowledge from home visiting and examine their impacts on multiple dimensions of skill in the early years.

We offer new insights into the measurement and evolution of skills and abilities. We empirically examine the microdynamics of child skill formation for multiple skills. We study the impact of the intervention on skill growth and present a nonparametric test of dynamic complementarity. We develop a new microdynamic model of multiple skill formation. It formalizes mechanisms suggested in developmental psychology.[1] We investigate the growth of skills at more granular levels than previous analyses.

We propose and implement tests for *age invariance* of skills, a crucial assumption maintained in the recent literature estimating technologies of skill formation. It as-

[1] See, e.g., Bronfenbrenner (2005); Thelen (2005).

sumes a constant-unit measuring stick for skills over ages and inputs.[2] Technology-based estimates of dynamic complementarity depend crucially on this scaling assumption, as do "value added" models for evaluating educational interventions.[3] We bypass this problem and test for dynamic complementarity without imposing any particular scale for skills.

Our point of departure is the *technology of skill formation* that characterizes the growth of child skills at age (stage) $\boldsymbol{a} : \boldsymbol{\theta_a}$. It is a function of a vector of family and governmental investment $\boldsymbol{I_a}$ (home visits, parenting, interactions with the child, school-based interventions, center care, school stimulation, etc.) and environments $\boldsymbol{G_a}$ (including neighborhood variables, parental education, and public goods):

$$\overbrace{\boldsymbol{\theta_{a+1}}}^{\substack{\text{skills at}\\ \boldsymbol{a+1}}} = \boldsymbol{f_a}\Big( \underbrace{\boldsymbol{\theta_a}}_{\substack{\text{Skills at}\\ a}}, \overbrace{\boldsymbol{I_a}}^{\text{Investment}}, \underbrace{\boldsymbol{G_a}}_{\substack{\text{Environmental}\\ \text{Variables}}} \Big). \tag{1}$$

Key properties featured in the literature are self productivity ($\boldsymbol{\theta}_{a+1} \uparrow \boldsymbol{\theta}_a$) and the productivity of investments and beneficial environments ($\boldsymbol{\theta}_{a+1} \uparrow \boldsymbol{I_a}, \boldsymbol{G_a}$). Dynamic complementarity ($\frac{\partial^2 \boldsymbol{\theta}_{a+1}}{\partial \boldsymbol{\theta}_a \partial \boldsymbol{I}'_{\boldsymbol{a}}} \geq 0 \uparrow a$) is a central proposition in the literature. It states that investment at one stage of the life cycle makes later investment more productive. It also implies that remediation of a skill deficit is more costly (requires

[2]For example, Todd and Wolpin (2007) and others use words spoken by age as a scale. The obvious question is whether twice as many words at age 5 is the same amount of knowledge as twice the same at age 8? Are % changes comparable at different ages? What is the appropriate metric? Are there common scales of knowledge? Is there a single-scale to measure growth of knowledge over time? For all skills? For any particular skill? An assumption of common scales ignores the multiple skills that emerge as a child matures and the nuance connected with the choice of words. In addition, many assessments bundle multiple skills (e.g., grades depend on cognitive and noncognitive skills) (Borghans et al., 2016). A growing body of evidence challenges the validity of psychometric conventions (see, e.g., Almlund et al., 2011 and Kautz et al., 2014).

[3]Cunha et al. (2010); Agostinelli and Wiswall (2021); Freyberger (2021).

more investment) the later the stage in the life cycle (Heckman and Mosso, 2014).

There are three big questions in this literature. (1) What is $\boldsymbol{I_a}$ and how to measure it? Many definitions are used.[4](2) What are the micro-mechanisms underlying the technology? Child psychologists emphasize parent/caregiver-child interactions— *"scaffolding"* (Vygotsky, 1978)—as major determinants of child development. They analyze the consequences of these interactions for learning and emotional development. Successful center-based programs also enrich children's home lives outside of childcare centers and keep parental engagement active long after children leave pre-K. Home visiting programs target family life directly (e.g., Elango et al., 2016).

(3) How should we measure skills and their growth? Test scores based on assessments of cognitive, socioemotional, and other skills are widely used.[5] Such measures have arbitrary scales (e.g., Uzgiris and Hunt, 1975; Cunha and Heckman, 2008; Cunha et al., 2010). Ordinal production functions that compare ranks across people at a point in time, but not growth in levels, do not suffer from this problem but, at the same time, do not measure levels of attained skill.[6] Dynamic complementarity increasing with age is intrinsically a *cardinal* property. Even if the measures used are monotone in knowledge, the property, $\frac{\partial^2 \boldsymbol{\theta_{a+1}}}{\partial \boldsymbol{\theta_a} \partial \boldsymbol{I_a'}} \uparrow a$ is *fundamentally cardinal*. One solution to this problem is to anchor scores (i.e., Cunha and Heckman, 2008) on meaningful outcomes (e.g., earnings, crime), but objective behavioral anchors at

[4]For example, books in the home, time spent in childcare/play, parenting styles (e.g., Doepke and Zilibotti, 2019; Kim, 2019; etc.), external interventions at centers or home visits. See, e.g., Cunha and Heckman (2008); Cunha et al. (2010); Del Boca et al. (2014); Agostinelli and Wiswall (2021); Andrew et al. (2019); Doepke and Zilibotti (2019).

[5]See, e.g., Kautz et al. 2014; OECD 2021.

[6]These remarks apply in full force to widely-used value added analyses in education as well.

early ages are difficult to construct.[7] Recent literature demonstrates the empirical importance of these issues. Freyberger (2021) shows the drastic consequences of different scalings for estimates of technology.

This paper addresses these questions using unique data. We report the following findings. First, there is strong evidence on the effectiveness of the home visiting interventions we study on child development measured in various ways. Impacts vary across skills, levels of skill, ages, family backgrounds, and child ability levels. Second, we develop new measures of learning, skills, and ability and examine the persistence of acquired skills. Different plausible measures of skill are not always strongly correlated with traditional test scores or each other. The metrics we develop can be fruitfully used to assess development. Third, we develop an empirically concordant dynamic model of learning. We find evidence supporting age invariance for certain skills at specific difficulty levels, but not globally for most skills. Fourth, without imposing age invariance, we find evidence for dynamic complementarity using non-parametric methods. We do this by showing the importance of previous learning in the acquisition of new skills, and by examining whether lagging children can catch up. Fifth, we determine the effects of ability and acquired skills on child learning rates for different skills and at different ages.

The plan of the paper is as follows. Section 2 describes the motivating program, Jamaica Home Visiting Intervention, and the program we analyze, China REACH, that is based on it. Section 3 discusses the data and the estimated conventional treatment effects. It develops and compares alternative measures of ability and growth of

[7] For a recent discussion of these problems, see Cunha et al. (2021).

knowledge and how they are impacted by various interventions. Section 4 presents nonparametric tests of dynamic complementarity. Section 5 develops microdynamic models of learning as latent Markov processes. We discuss identification of the model. Section 6 presents model estimates and their interpretation. Section 7 concludes.

## 2 The Program Analyzed and the Data

The inspiration for the program analyzed is the Jamaican Home Visiting Intervention (Grantham-McGregor and Smith, 2016). It was a randomized home visiting parenting intervention given to a sample of 129 stunted children between 9 and 24 months of age. It gave stunted children nutrition and also taught parents ways to interact with children. It is cheap and adaptable. Home visitors have about the same level of education as parents. The average cost is between $1,000 and $2,100 per child per year,[8] compared to more expensive omnibus programs like the Abecedarian program that can cost as much as $20,000 per child per year.

Substantial persistent effects are found for the program (i.e., Gertler et al., 2021, 2014). Appendix A summarizes the evidence on the substantial long-term effects of the Jamaica program through age 31 on education, IQ, social and emotional skills, and labor market outcomes.

Its success has spawned at least seven replications around the world, in Colombia, India, Bangladesh, Peru, and China. An Irish home visiting program was indirectly influenced by it. See Table A3 for a description of these programs.

The program we analyze is *China REACH*. It closely follows the Jamaican pro-

[8] 2021 USD.

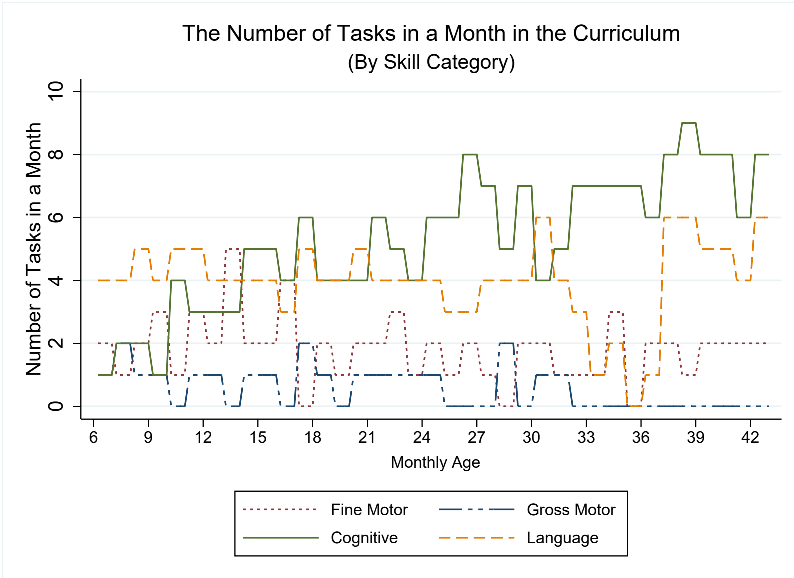
tocols. Implemented in 2015 by a large-scale random control trial, it enrolled 1,500 subjects (aged 6 months-42 months) in 111 villages in Huachi county, Gansu province, one of the poorest areas of China. Zhou, Heckman, Liu, and Lu (2021) improve on previous research by analyzing both conventional and difficulty-adjusted test scores. They report that estimated treatment effects derive from enhancing skills rather than from enhancing the use of existing skills.

China REACH is a paired-match RCT that achieves optimal mean square errors (Bai, 2019). A non-bipartite Mahalanobis matching method was used to pair villages and randomly select one village within the pair into the treatment group and the other village into the control group. Appendices B, C, and D describe the experiment, evidence on the balance of treatment and control groups, and numerous details on its implementation, including a study of attrition which turns out to be negligible. We briefly summarize its main features here.

Trained home visitors who are roughly at the program participants' mothers' level of education visit each treated household weekly and provide one hour of parenting or caregiving guidance. The curriculum teaches and encourages the mother/grandparent(s) to talk with the child through playing games, making toys, singing, reading, and storytelling to stimulate the child's cognitive, language, motor, and socioemotional skill development.

About 3 to 4 different skill tasks (gross motor, fine motor, language, and cognitive) are taught each week. Skills taught are ordered by difficulty level following profiles developed by Palmer (1971) and Uzgiris and Hunt (1975) and widely applied in the child development literature. Central to our identification strategy in this paper is

Figure 1: Curriculum Task Intensity by Skill Category




the assumption that these profiles describe valid hierarchies of knowledge. Child skills are assessed weekly. There are monthly assessments of the quality of visits by supervisors. Figure 1 presents the pattern of skill tasks taught and measured at each age.

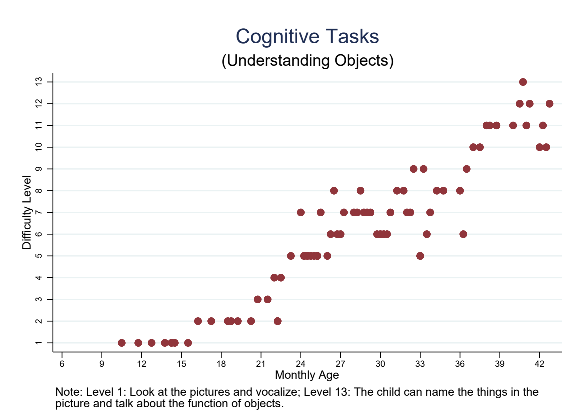
For cognitive skills, there are 13 difficulty levels. Table 1 gives the 13 difficulty levels for cognitive lessons and Figure 2 gives the timing of the lessons by age. Appendix E gives comparable information for the other skills.

The tasks advance cognitive skills from simply understanding a picture by verbal acknowledgment to using receptive (heard) language to identify pictures. Although the task content progresses by levels, the task content is similar within the same difficulty level. For example, the contents of cognitive skill tasks at level 1 are

Table 1: Difficulty Level List for Cognitive Lessons

| | |
|---|---|
| Level 1 | Look at the pictures and vocalize |
| Level 2 | Name the objects and ask the baby 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 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, 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 graphics, 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 and 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 2: The Timing of Cognitive Skill (Understand Objects) Tasks across Difficulty Levels



described in Table 2. All tasks relate to the activity of looking at pictures or objects and vocalizing.

Table 2: Cognitive Skill Task Content: Look at 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 |

Tasks evolve as depicted in Figure 2. As children age, they confront more demanding tasks. Occasionally, the protocol reverts to earlier levels of skill to review the child's learning and bolster confidence in their acquired skills. Reversion is relatively rare, and we ignore it in our empirical analysis.

Zhou, Heckman, Liu, and Lu (2021) use a modified Rasch-model analysis to estimate treatment effects. They report substantial treatment effects of the program and note that China REACH is on track to replicate or exceed the substantial outcomes of the Jamaica study.

The commonality of the protocols across multiple sites gives generality to the findings in this paper. Child development is a universal process. The goal of this paper is to investigate the sources of the reported treatment effects. We next present evidence on the effectiveness of the intervention and develop and compare a variety of measures of skill.

## 3 Measuring Skills and Their Evolution

We measure child development using records on weekly home visits. Supervisors record home visitor, caregiver, and child interaction activities at least once per month, making it possible for us to examine their impacts. We evaluate the interaction quality between home visitors and caregivers, and between home visitors and visited children. During the monthly home visit, program supervisors evaluate the home visit's quality in three dimensions: (a) Quality of the home visitor's teaching ability; (b) Interaction quality between the home visitor and the caregiver; and (c) Interaction quality between the home visitor and the child. Appendix F describes the interaction data and how we form factors to summarize it.

Zhou, Heckman, Liu, and Lu (2021) estimate the mean causal effects of interaction and home visitor quality on performance on standard achievement measures.[9] See Tables H2-H3. The estimated impacts of interactions between the home visitor and the caregiver are positive and statistically significant for cognitive and language skills. Estimated impacts of interactions between the home visitor and the child are generally not significant, nor is the teaching ability of the visitor. When we instrument for home-visitor interactions, we find stronger results.

Our data on weekly skill growth enable us to move beyond the traditional aggregates of % of items passed as reported in the Denver achievement tests to examine

[9]The Denver tests. See Appendix G.

age-by-age skill growth and factors that influence it. To understand the structure of our data, we introduce some notation. Let $\mathcal{S}$ be the set of skills taught. Let $\ell(s,a)$ be the level of skill $s$ taught at age $a$; $\ell(s,a) \in \{1,\ldots,L_s\}$. $L_s$ is the number of levels of difficulty for each skill $s$.

Mastery of skill $s$ at level $\ell$ at age $a$ is characterized by:

$$D(s,\ell,a) = \begin{cases} 1 & K(s,\ell,a) \geq \bar{K}(s,\ell) \\ 0 & \text{otherwise} \end{cases}$$

where $D(s,\ell,a)$ records mastery (or not) of a skill at a given level at age $a$. $\bar{K}(s,\ell)$ is the minimum latent skill required to master the task at difficulty level $\ell$. Let $\underline{a}(s,\ell)$ be the first age at which skill $s$ is taught at level $\ell$, and let $\bar{a}(s,\ell)$ be the last age at which it is taught at level $\ell$. For consecutive lessons in a run, $1+\bar{a}(\ell)-\underline{a}(\ell)$ is the length of run (# of lessons taught on skill $s$ at level $\ell$) starting at age $\underline{a}(s,\ell)$.

For level $\ell$ of skill $s$, collect the indicators of knowledge in a spell:

$$\Big\{D(s,\ell,a)\Big\}_{\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}.$$

In a stationary environment with age-invariant heterogeneity with no learning or growth of knowledge at level $\ell$, the sequences $\{D(s,\ell,a')\}$, $a' \in [\underline{a}(\ell),\bar{a}(\ell)]$ are exchangeable (i.e., they are equally probable for any order within $\ell$).[10]

With learning, sequences are back-loaded. For $j>0$,

$$\Pr(D(s,\ell,a+j) \geq D(s,\ell,a)) \geq \Pr(D(s,\ell,a+j) \leq D(s,\ell,a)).$$

[10]See Heckman (1978, 1981).

Knowledge acquisition for each skill $s$ at each level $\ell$ is measured by properties of these arrays and their relationships. No learning implies exchangeability by age for sequences within each $\ell$. Zhou, Heckman, Wang, and Liu (2021) test and reject this hypothesis for our data. We summarize our tests for exchangeability in Appendix I. We control for maturation and exposure effects that might boost skills in the absence of any intervention (see Appendix J). Even after doing so, we reject exchangeability and find evidence for knowledge growth.

Figure 3 characterizes the growth of knowledge in language, cognitive, and fine motor skills. Average passing rates within each difficulty level for language and cognitive tasks increase with age, a pattern consistent with learning. When individuals transition to a higher difficulty level, initial passing rates decline. Subsequent passing rates increase as learning ensues. Our dynamic model captures this phenomenon. The pattern for fine motor skills is somewhat different. For most levels, there is—at best—modest learning.[11] Access to detailed weekly data enables us to determine at what stages learning is taking place and which interventions are most effective. We next examine alternative measures of learning and learning speed and how they are related.

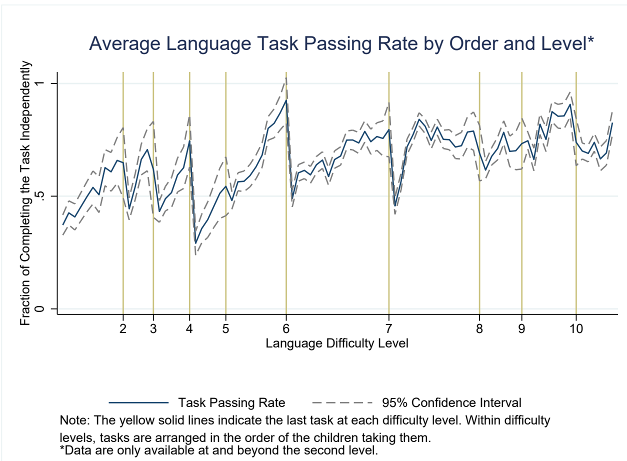
## 3.1 Measures of Knowledge and Knowledge Acquisition

The traditional measure of knowledge is the proportion of correct answers over all levels of difficulty within an assessment. The passing rate on skill $s$ at level $\ell$ is:

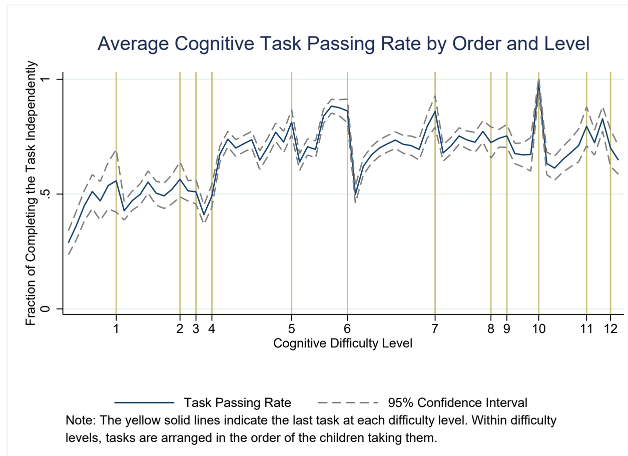
[11]We also measure gross motor skills, but they are very flat with age and are not affected by the intervention, so we do not systematically analyze them in the text.

Figure 3

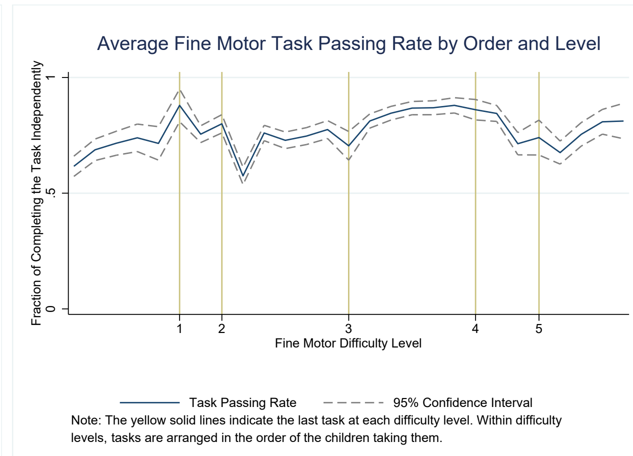
(a) Language




(b) Cognitive




(c) Fine Motor



$$p(s,\ell) = \frac{1}{\bar{a}(s,\ell) - \underline{a}(s,\ell) + 1} \sum_{a=\underline{a}(s,\ell)}^{\bar{a}(s,\ell)} D(s,\ell,a). \tag{2}$$

The overall passing rate is:

$$p(s) = \frac{\sum_{\ell=1}^{L_s} \Big\{1 + \bar{a}(s,\ell) - \underline{a}(s,\ell)\Big\} p(s,\ell)}{\sum_{\ell=1}^{L_s} \Big\{1 + \bar{a}(s,\ell) - \underline{a}(s,\ell)\Big\}}, \tag{3}$$

which weights all items across all difficulty levels equally, and puts more weight on difficulty levels with more items.

There are many other plausible measures of knowledge and knowledge acquisition. For consecutive learning spells with all participants entering each level at the first lesson, we define ***Time to first mastery*** as $d(s,\ell) = \hat{a}(s,\ell) - \underline{a}(s,\ell)$, where for each $s$ and $\ell$, $\hat{a}(s,\ell) = \min_a \{D(s,\ell,a) = 1\}_{a=\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}$. We can also define ***Time to full mastery*** as $\tilde{a}(s,\ell) = \min_a [D(s,\ell,a) = 1, \forall a \geq \tilde{a}(s,\ell)]$. These measures can also be averaged over all levels. Time to full mastery is $\tilde{a}(s,\ell) - \underline{a}(s,\ell)$. Other measures of learning are possible, such as time to mastery of two in a row after $\hat{a}(s,\ell)$, etc. We define ***Backsliding*** at level $\ell$ for skill $s$ as:

$$\frac{\#\{D(s,\ell,a) = 0, a > \hat{a}(s,\ell), a \leq \bar{a}(s,\ell)\}}{\#\{a > \hat{a}(s,\ell), a \leq \bar{a}(s,\ell)\}}.$$

It is instructive to examine the correlation between the traditional Denver achievement scores and these learning measures for each skill. Tables 3a-3c show the correlations between Denver scores (3 sub-scores: Language-cognitive, Fine Motor, and Gross Motor) and our learning measures by skill. All three of our measures are significantly correlated with the children's Denver Test Scores in the expected directions. The Denver Score is positively correlated with average passing rate across tasks during the intervention; the Denver score is negatively correlated with the time of the child achieving the first success; and the negatively correlated with the fraction of fails after the first success. We also find that compared to fine and gross motor scores, the language and cognitive scores have more statistically significant correlations with these three measures. The program significantly improves language and cognitive skills.

Table 3a: Correlation between Average Passing Rate (All Levels) and Denver Endline Scores

| | | Average Passing Rate | | | |
|---|---|---|---|---|---|
| | | Language | Cognitive | Fine Motor | Gross Motor |
| Denver Score (Endline) | Language and Cognitive | 0.078*** | 0.098*** | 0.099*** | 0.058*** |
| | Fine Motor | 0.011 | 0.042*** | 0.042** | 0.017 |
| | Gross Motor | 0.075*** | 0.088*** | 0.064*** | 0.055*** |

1. Average Passing Rate is the passing rate for the intervention tasks at each difficulty level by each skill type.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3b: Correlation between Time to Mastery (All Levels) and Denver Endline Scores

| | | Time to Mastery | | | |
|---|---|---|---|---|---|
| | | Language | Cognitive | Fine Motor | Gross Motor |
| Denver Score (Endline) | Language and Cognitive | -0.077*** | -0.071*** | -0.046** | 0.018 |
| | Fine Motor | -0.027 | -0.028* | -0.012 | -0.005 |
| | Gross Motor | -0.069*** | -0.068*** | -0.016 | -0.025 |

1. Time to mastery is defined as the number of tasks a child takes until the first success (inclusive) at each difficulty level during the intervention by each skill type.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3c: Correlation between Instability (All Levels) and Denver Endline Scores

| | | Instability | | | |
|---|---|---|---|---|---|
| | | Language | Cognitive | Fine Motor | Gross Motor |
| Denver Score (Endline) | Language and Cognitive | -0.070*** | -0.063*** | -0.043* | -0.078*** |
| | Fine Motor | -0.026 | -0.040** | -0.021 | -0.031 |
| | Gross Motor | -0.061*** | -0.074*** | -0.048** | -0.061** |

1. Instability is defined as fraction of fails after the first success at each difficulty level by each skill type.
2. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

While all the correlations are in the expected direction, the different measures are far from perfectly correlated, suggesting that they capture different dimensions of knowledge.[12] Instability (Backsliding) is at best weakly correlated with speed (time to mastery).

Table 4: Correlations Among the Three Measures of Learning by Skill

| Correlation Variables | Language | Cognitive | Fine Motor | Gross Motor |
|---|---|---|---|---|
| Time to Mastery vs. Avg. Passing Rate | -0.641*** | -0.677*** | -0.688*** | -0.607*** |
| Time to Mastery vs. Instability | 0.181*** | 0.208*** | 0.175*** | -0.035 |
| Avg. Passing Rate vs. Instability | -0.810*** | -0.831*** | -0.857*** | -0.932*** |

*Notes:* Average Passing Rate is the passing rate for the intervention tasks at each difficulty level by each skill type. 2. For intervention tasks, instability is defined as proportion of fails after the first success at each difficulty level by each skill type. 3. Time to Mastery is defined as the number of tasks a child takes until the first success (inclusive) at each difficulty level.
4. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

Appendix K summarizes the dimensionality of these measures. There are two dimensions for each skill and at least five dimensions across all skills. The notion of a single dimension of skill assumed in standard efficiency unit models and in the psychology of "g" that claims one universal skill predicts performance on all tasks, is grossly inaccurate.

[12] An alternative explanation is substantial measurement error. Our factor analyses of these data show that measurement error ("uniqueness") is a real possibility. See Cunha et al. (2021) for a discussion of measurement error in such measures.

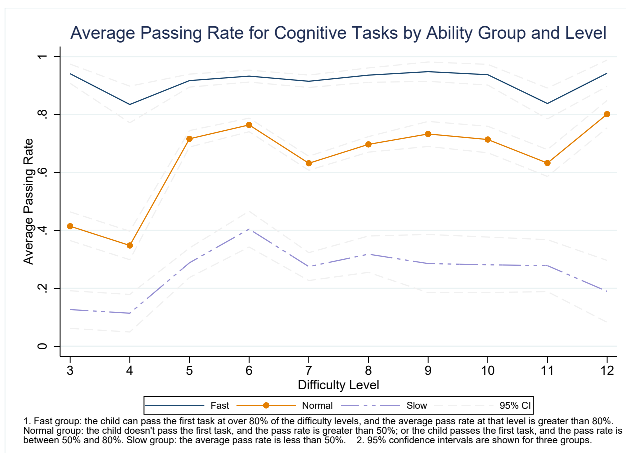
Table 5: Ability Groups

| | |
|---|---|
| **Fast group** | Pass the first task for more than 80% of difficulty levels, and the average passing rate of all the skill specific tasks is more than 80%. |
| **Normal group** | Pass the first task for less than 80% of difficulty levels, and the average passing rate of all the skill-specific tasks is between 50% and 80%. |
| **Slow group** | The average passing rate of all the skill-specific tasks is less than 50%. |

Using our data, we can define ability groups and determine their stability of membership in categories. The categories are defined by performance across levels on the designated skills.

There is strong persistence of passing rates across difficulty levels (see Figure 4).

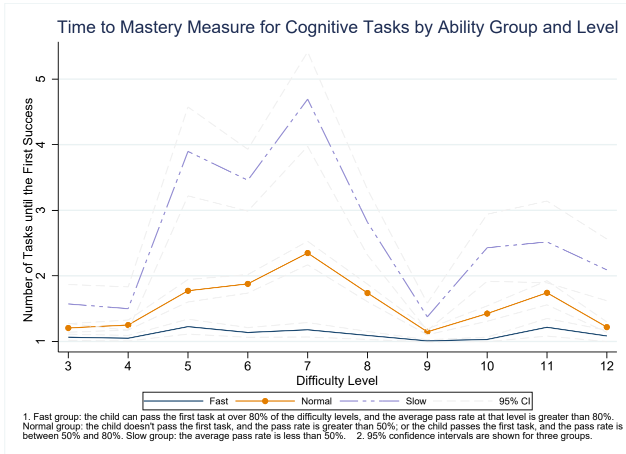
Figure 4: Average Passing Rate for Cognitive Tasks by Ability Group and Level




*Notes:* For cognitive skills: the fast group represents 27.53%, the normal group represents 57.4%, and the slow group represents 15.09%.

Figure 5 quantifies the persistence of the speed (# of tasks required to achieve mastery) across difficulty levels for cognitive skills. Figure L2 displays a similar pattern for all skills.

Figure 5: Time to Mastery for Cognitive Tasks by Ability Group and Level




Ability predicts the fraction of times that children get the wrong answer after a first correct answer (a measure of instability in performance) for cognition, language, and the other skills. See Figure L3.

## 3.2 Impacts of Interventions on Skill Levels

Using these measures, we apply the model used to estimate program treatment effects. We control for endogeneity of visitor quality in the fashion described in Appendix H. We use the information on the performance in terms of child outcomes of the same visitor in different spatially separated villages to form instruments.

Table 6 reports the impact of program interactions on time to the first mastery of achieving language tasks. It shows a recurrent pattern. The interaction between the home visitor and the caregiver is the only consistently statistically significant pattern across all difficulty levels (note that a negative coefficient means a quicker mastery).

Note further that age (maturation) effects are statistically important. Children acquire skills with age and experience. Having a grandmother as the caregiver has negative effects on achieving the task goal fast at higher levels, i.e., retards learning speed.

Table 6: The Effects of Interactions on the Time to First Mastery of Achieving Language Tasks at Each Level (IV)

| | Language Task Difficulty Levels | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\leq 2$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| Interaction Quality | $-1.893^{**}$ | $-0.729^{***}$ | $-0.868^{***}$ | $-0.252$ | $-0.974^{**}$ | $-0.926^{**}$ | $-0.365$ | $-0.263^{*}$ | $-0.265^{*}$ | $-0.298^{*}$ |
| Home Visitor and Caregiver | (0.774) | (0.235) | (0.321) | (0.632) | (0.380) | (0.389) | (0.270) | (0.142) | (0.142) | (0.170) |
| Interaction Quality | $-0.292$ | $-0.053$ | $-0.154$ | $-0.163$ | $-0.351^{**}$ | $-0.278^{*}$ | $-0.138$ | 0.060 | $-0.045$ | 0.019 |
| Home Visitor and Child | (0.375) | (0.098) | (0.149) | (0.239) | (0.154) | (0.161) | (0.106) | (0.056) | (0.068) | (0.045) |
| Teaching Ability | $1.767^{***}$ | $0.684^{***}$ | $0.692^{**}$ | 0.927 | $1.619^{***}$ | $1.230^{**}$ | -0.326 | $-0.470^{**}$ | $0.264^{**}$ | 0.341 |
| | (0.591) | (0.231) | (0.309) | (0.602) | (0.355) | (0.507) | (0.376) | (0.239) | (0.128) | (0.230) |
| Grandmother Rearing[1] | 0.970 | 0.070 | $-0.131$ | 0.836 | 0.218 | $0.993^{*}$ | 0.773 | $0.744^{***}$ | $1.081^{***}$ | $0.540^{*}$ |
| | (1.520) | (0.343) | (0.481) | (0.741) | (0.582) | (0.551) | (0.472) | (0.273) | (0.287) | (0.286) |
| Monthly Age | $-0.114^{***}$ | $-0.048^{**}$ | $-0.093^{***}$ | $-0.097^{***}$ | $-0.074^{***}$ | $-0.036^{**}$ | 0.014 | $-0.025^{*}$ | 0.004 | $-0.036^{*}$ |
| | (0.020) | (0.020) | (0.016) | (0.033) | (0.014) | (0.014) | (0.022) | (0.013) | (0.013) | (0.020) |
| Constant | $4.574^{***}$ | $2.246^{***}$ | $3.350^{***}$ | $3.370^{***}$ | $3.348^{***}$ | $2.912^{***}$ | $1.936^{***}$ | $2.125^{***}$ | $1.035^{**}$ | $2.286^{***}$ |
| | (0.462) | (0.330) | (0.281) | (0.505) | (0.288) | (0.384) | (0.518) | (0.351) | (0.465) | (0.542) |
| Observations | 206 | 148 | 178 | 92 | 374 | 485 | 490 | 375 | 308 | 432 |

1. % of home visits when grandmother is the primary caregiver. 2. The estimates reported in the table are based on the instrumental variable regression. 3. The variables of teaching ability, and interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. 4. The instrumental variables include mean, max, and min of other village interaction measures through the same visitor. 5. Time to Mastery is defined as the number of tasks a child takes until the first success (inclusive) at each difficulty level. 6. Standard errors in parentheses are clustered at village level. 7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix M reports comparable results for other skills. The interventions have no effects on gross motor skills.

## 3.3 Impact of the Interventions on Post-Treatment Interaction with Children

The program's goal is to promote interactions in play that are designed to promote learning between the caregiver and the child. We lack detailed information on these interactions. We measure the frequency of the caregiver playing with the child on the tasks after each home visit.[13] We analyze the average over all visits. Table 7 shows that the instrumented intervention variables only promote play for low-ability children, where ability is measured by simultaneous membership in the ability groupings for cognitive, language, and fine motor skills. Grandmothers are less likely to play with children.

Table 7: The Effects of Interactions on the Frequency of the Caregiver Playing with the Child by Ability Group

| | Fast | | Slow | |
|---|---|---|---|---|
| | OLS | IV | OLS | IV |
| Interaction Quality: | 0.030 | 0.059 | 0.154 | 0.376*** |
| Home Visitor and Caregiver | (0.238) | (0.507) | (0.091) | (0.073) |
| Interaction Quality: | 0.207* | 0.059 | -0.036 | -0.108*** |
| Home Visitor and Child | (0.114) | (0.108) | (0.020) | (0.019) |
| Teaching Ability | 0.188 | 0.056 | 0.077 | -0.041 |
| | (0.197) | (0.266) | (0.073) | (0.168) |
| Grandmother Rearing | -0.408 | -0.053 | -0.062 | -0.145** |
| | (0.415) | (0.489) | (0.093) | (0.063) |
| Monthly Age | -0.000 | -0.002 | 0.001 | -0.002 |
| | (0.012) | (0.014) | (0.004) | (0.003) |
| Constant | 5.194*** | 5.500*** | 5.745*** | 5.842*** |
| | (0.456) | (0.608) | (0.144) | (0.148) |

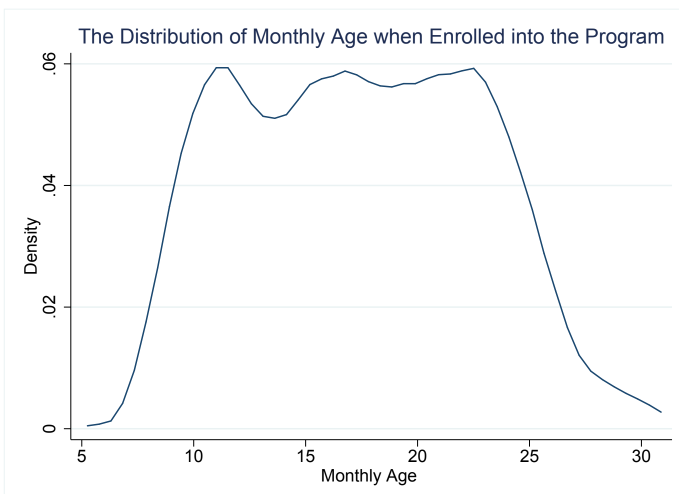
1. Frequency of the caregiver playing with the child is defined as: the number of days in a week that the caregiver plays with the child using tasks from the last home visit.
2. The fast group in the table is defined as the children who are in the fast group for language, cognitive, and fine motor skills. Similarly, the slow group is defined as the children who are in the slow group for language, cognitive, and fine motor skills, where the previous definition of fast and slow are used.
3. The instrumental variables include mean, max, and min of other village interaction measures using the same visitor.
4. Standard errors in parentheses are clustered at village level.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[13] Specifically, we record the following information: the number of days in a week that the caregiver plays with the child using tasks from the last home visit.

# 4 Nonparametric Tests of Dynamic Complementarity

Before turning to our dynamic model of skill growth, we present some nonparametric evidence on dynamic complementarity, which does not require any particular assumptions on scales of skills. Although children enter the program at different ages, all enrolled children of the same age receive the same lesson. We determine whether late arrivals can catch up. This is an aspect of dynamic complementarity: how rapidly do children who enter the program later improve their skills compared to those who entered earlier and had some skill training.

Figure 6: The Distribution of Monthly Age when Enrolled into the Program




Over the age range 10-23 months, children enter the program more or less randomly with respect to age due to administrative constraints (see Figure 6). None of the children receive training in the program before entry but may acquire skills from imitation and maturation. Suppose that a child enters at level $\ell(s)$ at skill $s$ at age $a^+(s,\ell)$. Some may be able to master the task from the outset due to maturation and exposure to rich environments, but many do not. We compute the probability of mastery at entry age $a^+(s,\ell)$ for the new entrants as

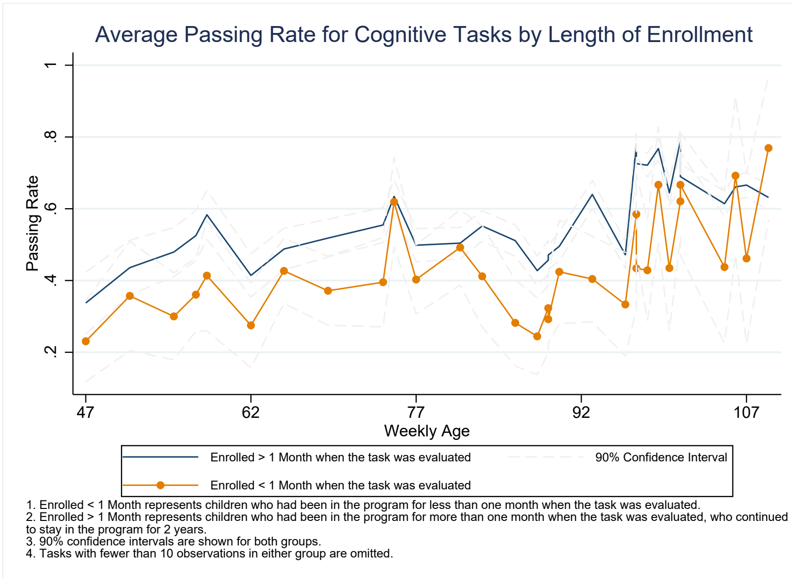
$$q(s,\ell,a^+(s,\ell)) = \Pr\Big(D(s,\ell,a^+)=1\Big) \text{ for } a \geq a^+(s,\ell).$$ [14]

It is a measure of learning from maturation and exposure without participating in the program. It also informs us about dynamic complementarity and the importance of initial conditions.

We consider performance by age at the entry level. Figure 7 shows the initial passing rate ($q$) for cognitive tasks by age (length of enrollment). We find that for most tasks, the group that is enrolled for longer than one month performs significantly better than the new entrant group.

[14] We define new entrants as children who enroll in the program and have less than one month of exposure to it.

Figure 7: Cognitive Tasks Performance Comparison by Length of Enrollment




Appendix N presents comparable figures (N1–N16) for other skills. The pattern is the same except for gross motor skills, which appear to be little affected by participation in the program.

The longer the child has been in the program, the higher the passing rate on cognitive tasks. Figure 8a shows that this pattern dissipates by the end of the program, when there is apparent catch-up. This pattern is consistent with skill depreciation across levels or the acquisition of new skills at later levels.[15] A comparable pattern appears in the figures for other skills in Appendix N.

Of course, difficulty may vary across levels, so we must be careful in inferring too much from these graphs. Estimates from our structural model clarify this evidence.

[15]Students near the end of the program have been in it for at least 6 months. They are effectively not new entrants (see Figure 6). Six months may be enough to master the skills examined at later ages, but this is an interesting finding in itself.

Figure 8a: Average Passing Rate for Language Tasks by Enrollment Age

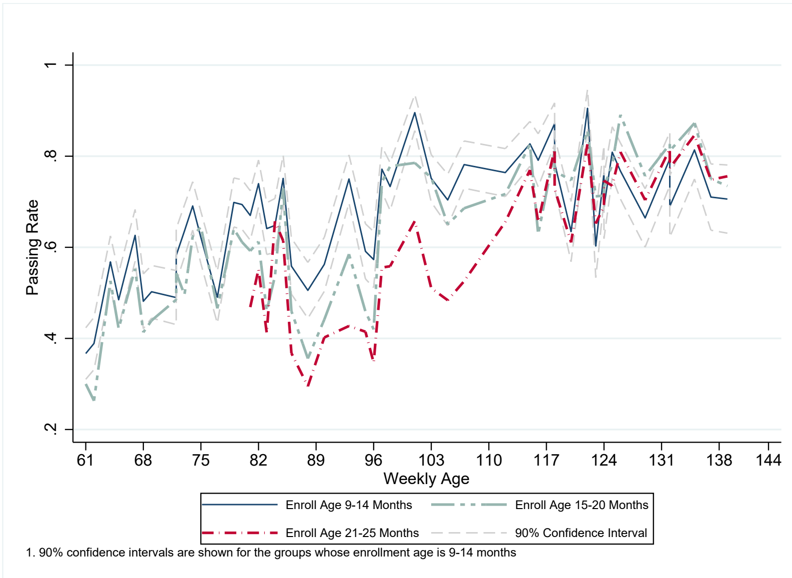



Figure 8b: Average Passing Rate for Language Tasks by Enrollment Age and Ability Group

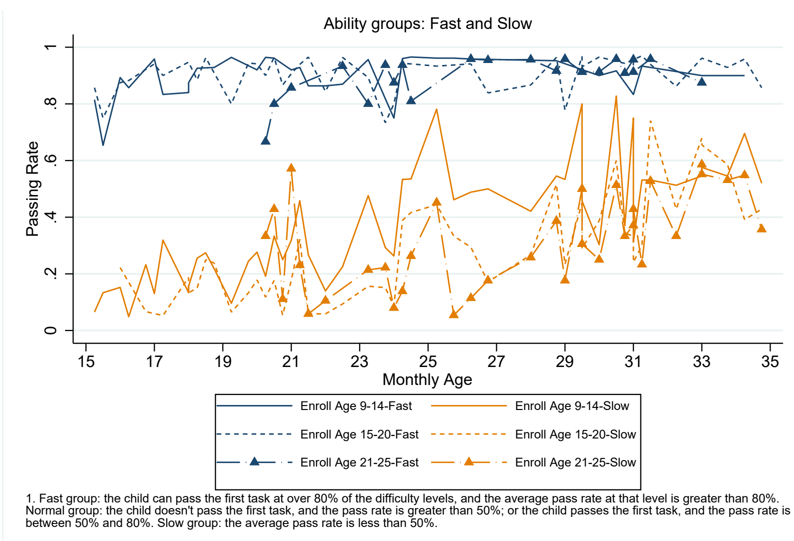


Figure 8c: Average Passing Rate for Language Tasks by Enrollment Age and Ability Group

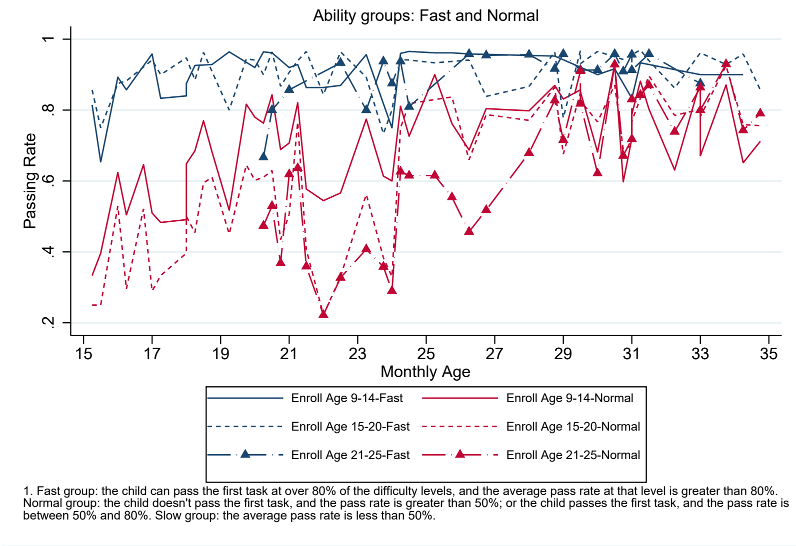
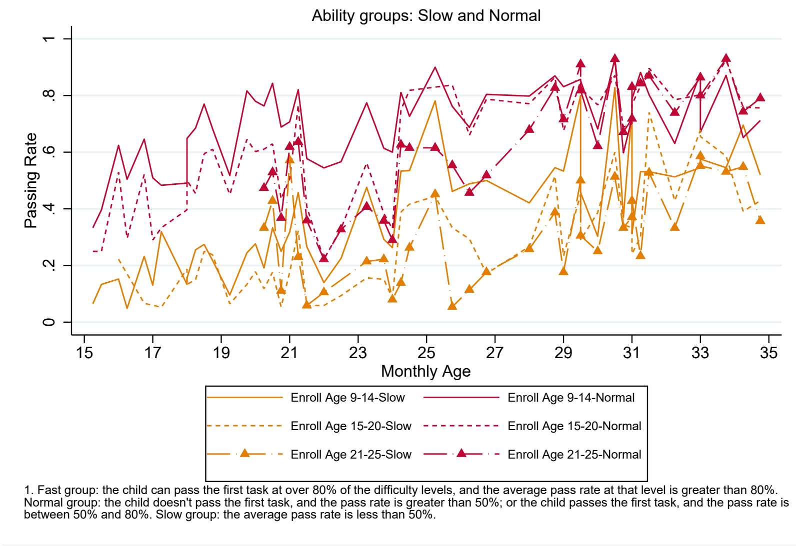



Figure 8d: Average Passing Rate for Language Tasks by Enrollment Age and Ability Group



The pattern of catch up varies by ability group. High ability children starting late catch up in language skills almost immediately (see Figure 8b). Low ability children who start late, catch up to low ability children who start early only near the end. Low ability children never catch up to high ability children. Normal children who start late catch up to normal children who start early only near the end. They also nearly catch up with fast children (see Figure 8c). Slow children who start late catch up to low children who start early, but only near the end (see Figure 8d) and slow children never catch up to normal children. The rapid learning of "fast" children suggests that an adaptive curriculum may be more beneficial for them. They can be boosted to more challenging levels more quickly than average and slow children, and learning would be promoted.

Similar patterns emerge for learning cognitive tasks (see Appendix Figures N6–N10), fine motor tasks (see Appendix Figures N11–N15), and gross motor tasks (see Appendix Figures N16–N20).

# 5 Mechanisms Generating Child Learning

To motivate our approach to estimating the weekly dynamics of skill formation, we consider a simple model for one level of skill before presenting our general model. The more general model is the simple model applied to each skill at each level. The program fosters skill at ages $a \in [0, \ldots, \bar{A}]$. Lessons are the same for everyone at age $a$. Define $K(a)$ as the level of "knowledge" at age $a$ with the initial value $K(0)$.

Lessons with identical skill content are taught and examined using a series of tasks. A person exhibits *mastery* of a skill at level $\bar{K}$ if $K(a) \geq \bar{K}$. Let $D(a) = 1$ if the person at age $a$ masters the skill, so $D(a) = \mathbf{1}\big(K(a) \geq \bar{K}\big)$. Mastery is measured at each age.

Consider a deterministic model of skill formation. Assuming no depreciation, skill evolves via

$$K(a) = K(a-1) + \delta(a)\eta K(a-1) + V(Q(a)), \tag{4}$$

where $\eta$ is the ability to learn parameter that is individual specific and assumed positive ($\eta > 0$), and $\delta(a)$ is the "lesson" at age $a$ for everyone enrolled. $V(Q(a))$ captures variables $Q(a)$ that affect the evolution of skills that operate independent of the level of $K(a-1)$. We assume scale invariance *within* each skill level. Thus skills are additive in the metric that quantifies $K$.

In this framework, *Self-Productivity* is $\dfrac{\partial K(a)}{\partial K(a-1)} = 1 + \delta(a)\eta$. *Investment Productivity* is $\dfrac{\partial K(a)}{\partial \delta(a)} = \eta K(a-1)$. *Dynamic Complementarity* is captured by investment $\delta(a)$ being more productive the higher the skill level:

$$\frac{\partial^2 K(a)}{\partial K(a-1)\partial \delta(a)} = \eta > 0.$$

Ability $\eta$ is a determinant of dynamic complementarity and boosts self-productivity. Ability itself can be a consequence of investment and operates like $K(a-1)$ in boosting skills.[16]

[16] If there is geometric depreciation, $K(a) = (1-\sigma)K(a-1) + \delta(a)\eta K(a-1) + V(Q(a))$. Self-productivity is $1 - \sigma + \delta(a)\eta$. Investment productivity is as before, as is dynamic complementarity.

## Adding Shocks

A multiplicative version of the model turns out to fit the data on skill growth very well. Adding i.i.d. idiosyncratic shocks in growth rates ($\varepsilon(a)$) on a log scale, skill acquisition is characterized by:

$$\ln K(a) - \ln K(a-1) \doteq \delta(a)\eta + V(Q(a)) + \varepsilon(a)^{17} \tag{5}$$

so

$$\ln K(a) \doteq \eta \sum_{j=1}^{a} \delta(j) + \sum_{j=1}^{a} V(Q(j)) + \sum_{j=1}^{a} \varepsilon(j) + \ln K(0) \tag{6}$$

where $\varepsilon(a)$ is i.i.d. across all $a$ with $E(\varepsilon(a)) = 0$. The model exhibits dynamic complementarity, self-productivity, and investment productivity. It introduces random walk growth in knowledge following Rutherford (1955).

Define $U(a) = \sum_{j=1}^{a} \varepsilon(j)$ a random walk, $\Delta(a) = \sum_{j=1}^{a} \delta(j)$, cumulative lessons, and $\Lambda(a) = \sum_{j=1}^{a} V(Q(j))$. The probability of mastery of the skill at age $a$ is

$$\Pr(D(a) = 1) = \Pr(\ln K(0) + U(a) + \Lambda(a) + \eta\Delta(a) > \ln \bar{K}),$$

where we assume $\eta \perp\!\!\!\perp \varepsilon(j)$ for all $j$. Thus, conditioning on $\eta$, assumed to be independent of $U(a)$, $\eta$ and $K(0)$, we obtain

[17] With geometric depreciation, we replace $\delta(a)\eta$ with $\delta(a)\eta - \sigma$.

$$\Pr(D(a) = 1 \mid \eta, \Delta(a), \Lambda(a), K(0)) = \int_{\ln \bar{K} - \eta\Delta(a) - \Lambda(a) - \ln K(0)}^{\infty} dF(U(a)). \qquad (7)$$

## The General Model

Using notation introduced in Section 3, the general model is the same as the simple model applied to skills at each level where $\mathcal{S}$ is the set of skills taught, $\ell(s, a)$ is the level of skill $s$ taught at age $a$, and $\ell(s, a) \in \{1, \ldots, L_s\}$, where $L_s$ is the number of levels of difficulty for each skill $s$.

Following traditions in psychometrics, we use a threshold crossing model to characterize mastery of skill $s$ at level $\ell$ when the agent is $a$ years old.[18]

$$D(s, \ell, a) = \begin{cases} 1 & K(s, \ell, a) \geq \bar{K}(s, \ell) \\ 0 & \text{otherwise.} \end{cases}$$

$\bar{K}(s, \ell)$ is the minimum latent skill requirement to pass the task for difficulty level $\ell$. Shocks at level $\ell$ for age $a$—$\varepsilon(s, \ell, a)$—are assumed to be independent across $a$. Its distribution may vary with $\ell$ and $a$, but we later assume it is i.i.d. within $\ell$. $\eta(s)$ may vary across $\ell$ (ability may be level-specific) and $\delta(a)$ might differ as lessons change. Thresholds (passing standards) $\bar{K}(s, \ell)$ may also change across levels, as may $V_\ell(Q(a))$. We next define skill invariance and show how we test for it within our model.

[18] See, e.g., Lord and Novick, 1968 or van der Linden, 2016.

## 5.1 Skill Invariance

Skill invariance assumes a common scale within and across all $\ell$ for each $s$, although scales may vary across $s$. Index $K(s, \ell, a)$ cumulates over levels so the notion of knowledge growth is well-defined. It requires, among other things, that in the absence of depreciation

$$\underbrace{K(s, \ell, \underline{a}(s, \ell))}_{\text{Initial condition at level } \ell} = \underbrace{K(s, \ell - 1, \bar{a}(s, \ell - 1))}_{\text{Terminal condition at level } \ell-1} .$$

If all components of the technology of skill formation (Equation (5)) shift across levels, the notion of age-invariant scales lacks testability because the scale is not directly observed, and technology parameters can be redefined to impose age invariance. Some parameters must be invariant across levels, although they need not necessarily be the same parameters across all levels. Our proof of model identification in Appendix O makes this point precise.

If there is depreciation (or appreciation) by level,

$$K(s, \ell, \underline{a}(s, \ell)) = \Gamma_\ell(K(s, \ell - 1, \bar{a}(s, \ell - 1))),$$

where $\Gamma_\ell$ is a general function. Total depreciation of skills at $\ell - 1$ sets $\Gamma_\ell$ to the zero function. Age invariance at $\ell - 1$ sets $\Gamma_\ell = I$, the identity function. In this paper, we consider only affine transformations

$$\Gamma_\ell(K(s, \ell, \underline{a}(s, \ell))) = \gamma_{0,\ell} + \gamma_{1,\ell}(K(s, \ell - 1, \bar{a}(s, \ell - 1))).$$

We next discuss the intuition for model identification. We do not have direct measures of latent skills, but instead have strings of binary task performances for children enrolled in the program, from which we can infer their skills up to scale as in the standard binary threshold crossing model (see, e.g., Matzkin, 1992).

## 5.2 Identification

In order to avoid notational complexity, we use a simplified notation for a single skill to motivate essential ideas underlying model identification. A formal proof is presented in Appendix O. We focus on means and covariances because we assume normal errors in estimation. Drawing on Heckman and Vytlacil (2007) and Matzkin (1992, 2007), we can nonparametrically identify the joint distributions up to normalization under conditions stated in those papers. Define the latent index $K(1, a)$ for skill at level 1 at age $a$. This corresponds to $K(s, 1, a)$ for a particular skill $s$, which is kept implicit. We simplify Equation (5) to read:

$$\ln K(1, a) = \underbrace{V_1(a)}_{\text{maturation effects}} + \underbrace{\delta_1(a)\eta}_{\text{learning}} + \underbrace{U(1, a)}_{\text{shocks}} + \ln K(0), \tag{8}$$

where $K(1, a)$ is the latent index (skill) of a binary outcome model at difficulty level 1 at weekly age $a$, and $K(0)$ is the initial condition. $\ln K(0) = \mu_0(Z) + \Upsilon$, where $Z$ are background variables, $E(\Upsilon) = 0$, $\Upsilon \perp\!\!\!\perp \eta$, and $Z \perp\!\!\!\perp \Upsilon$. $U(1, a) = \sum_{j=1}^{a} \varepsilon(1, j)$, where $\varepsilon(1, j)$ is a task-specific shock at difficulty level 1 at weekly age $j$, which is assumed to be i.i.d. with variance $\sigma^2_{\varepsilon(1)}$. We assume that $\varepsilon(1, j) \perp\!\!\!\perp (\eta, \Upsilon)$ for all $j$. We parameterize $\delta_1(a)\eta = \bar{\beta}_1(X) + \omega_1$, where $X$ are covariates explaining ability and

$X \perp\!\!\!\perp \varepsilon(1,j)$ for all $j$. $V_1(a)$ is shorthand for $V_1(Q(a))$. We assume that the learning component is constant at each level but may differ across levels. In this notation,

$$D(1,a) = 1(\ln K(1,a) > \bar{K}(1)).$$

In general, we cannot distinguish thresholds apart from any intercepts associated with the technology and the initial condition.

A random coefficient model for ability multiplied by lesson content is $\beta_1 = \bar{\beta}_1(X) + \omega_1$, where $\omega_1$ is a level-specific random shock, $E(\omega_1) = 0$, $\omega_1 \perp\!\!\!\perp X$, and $\omega_1 \perp\!\!\!\perp (\Upsilon, \varepsilon(1,j))$ for all $j$. The random coefficient captures heterogeneity in learning ability. Equation (8) can be rewritten in this notation for the general case allowing for heterogeneity in $\ln K(0)$:

$$\ln K(1,a) = \mu_1 + \mu_0(Z) + V(Q(a)) + \bar{\beta}_1(X)a + \underbrace{\left\{ a\omega_1 + \sum_{j=1}^{a} \varepsilon(1,j) + \Upsilon \right\}}_{\Psi_1(a)} \tag{9}$$

where $\mathrm{Var}(\Psi_1(a)) = a^2\sigma^2_{\omega_1} + a\sigma^2_{\varepsilon(1)} + \sigma^2_{\Upsilon} := \sigma^2(1,a)$, where $\sigma^2(1,1) = \sigma^2_{\omega_1} + \sigma^2_{\varepsilon(1)} + \sigma^2_{\Upsilon}$.

Under conditions given in Matzkin (1992, 2007), with sufficient variation in the regressors for this threshold crossing model, we can identify in period $j$, $\underline{a}(1) \leq j \leq \bar{a}(1)$,

$$\frac{\mu_1^*}{\sigma(1,j)}, \quad \frac{\mu_0(Z)}{\sigma(1,j)}, \quad \frac{\bar{\beta}_1(X)}{\sigma(1,j)}, \quad \frac{V_1(Q(a))}{\sigma(1,j)},$$

where $\mu_1^* = \mu_1 - \bar{K}(1)$ and $\mu_1$ collects any other model intercepts. If any coefficient is common across $j$ and $j'$, we can identify the ratio of $\frac{\sigma(1,j)}{\sigma(1,j')}$. Under this condition, with one normalization (e.g., $\sigma(1,j)=1$), we can identify $\mu_1^*$, $\mu_0(Z)$, $\bar{\beta}_1(X)$, $V(Q(a))$ up to scale.

From the covariance between $D(1,a)$ and $D(1,a')$, we can identify

$$\mathrm{Cov}\left(\frac{\Psi_1(a)}{\sigma(1,a)},\frac{\Psi_1(a')}{\sigma(1,a')}\right)=\left\{(aa')\sigma^2_{\omega_1}+\min(a,a')\sigma^2_{\varepsilon(1)}+\sigma^2_{\Upsilon}\right\}\frac{1}{\sigma(1,a)\sigma(1,a')},$$

where $\sigma(1,a)$, $\sigma(1,a')$ are identified up to a normalization (e.g., $\sigma(1,1)$) (see Heckman, 1981 and Heckman and Vytlacil, 2007). For $a\geq 2$, we can thus identify $\sigma^2_{\omega_1}$, $\sigma^2_{\varepsilon(1)}$, and $\sigma^2_{\Upsilon}$ up to scale within each $\ell$ using age variation in the term in braces.

To simplify the analysis, we assume that $\omega_\ell=\omega$ for $\ell\in\{1,\ldots,L\}$. We can relax this assumption and still achieve identification. However, we have to take a position on the dependence across $\omega_j$.[19]

Adopting a similar notation for levels $\ell>1$, if we assume age-invariant measures connecting level 1 with level 2 (i.e., $\gamma_{0,2}=0$, and $\gamma_{1,2}=1$), we can connect latent skill $\ln K(1,\bar{a}(1))$ (the index of the last age $\bar{a}(1)$ of the last task at level 1) to the initial skill at level 2, $\ln K(2,\underline{a}(2))$: $\ln K(1,\bar{a}(1))=\ln K(2,\underline{a}(2))$. The latent skill at

[19]One attractive alternative assumption that secures identification is $\omega_j=\rho\omega_{j-1}+\tau_j$, where $\tau_j$ is mean zero, i.i.d over $j$.

level 2 at age $a$ can be written as:

$$
\begin{aligned}
\ln K(2,a) &= \mu_2 + V_2(Q(a)) + \bar{\beta}_2(X)(a-\bar{a}(1)) + \sum_{j=\underline{(a)}(2)}^{a} \varepsilon(2,j) + \ln K(1,\bar{a}(1)) \\
&= \mu_1 + \mu_2 + \mu_0(Z) + V_1(Q(\bar{a}(1))) + V_2(Q(a)) + \bar{\beta}_2(X)(a-\bar{a}(1)) + \bar{\beta}_1(X)\bar{a}(1) \\
&+ \underbrace{\left\{ \sum_{j=\underline{a}(2)}^{a} \varepsilon(2,j) + (a-\bar{a}(1))\omega + \sum_{j=1}^{\bar{a}(1)} \varepsilon(1,j) + \bar{a}(1)\omega + \Upsilon \right\}}_{\Psi_2(a)}. \qquad (10)
\end{aligned}
$$

Given the initial normalization and the identification of the parameters in the first level (up to scale), we can identify $V_2(Q(a))$ and $\bar{\beta}_2(X)$ up to scale $\sigma(2,a)$, where

$$
\Psi_2(a) = \sum_{j=\underline{a}(2)}^{a} \varepsilon(2,j) + (a-\bar{a}(1))\omega + \sum_{j=1}^{\bar{a}(1)} \varepsilon(1,j) + \bar{a}(1)\omega + \Upsilon.
$$

$$
\begin{aligned}
\sigma^2(2,a) :=& \mathrm{Var}\Psi_2(a). \\
\mathrm{Var}\Psi_2(a) =& \sigma^2_\Upsilon + a^2\sigma^2_\omega + (a-\bar{a}(1))\sigma^2_{\varepsilon(2)} + \bar{a}(1)\sigma^2_{\varepsilon(1)}.
\end{aligned}
$$

$$
\mathrm{Cov}\left(\frac{\Psi_2(a)}{\sigma(2,a)}, \frac{\Psi_2(a')}{\sigma(2,a')}\right) = \frac{\sigma^2_\Upsilon + aa'\sigma^2_\omega + (\bar{a}(1)-\underline{a}(1))\sigma^2_{\varepsilon(1)} + \min((a-\underline{a}(2)),(a'-\underline{a}(2)))\sigma^2_{\varepsilon(2)}}{\sigma(2,a)\sigma(2,a')}.
$$

We can identify $\sigma^2_{\varepsilon(2)}$ since all of the other terms are identified by the previous analysis. By an analysis similar to that for level 1, we can identify $V_2(Q(a))$ and $\bar{\beta}_2(X)$ up to $\sigma(2,a)$. From coefficients common across levels, we can identify $\sigma(2,a)$ for $\underline{a}(2,a) \le a \le \bar{a}(2,a)$.

Note further that across levels, we can identify (recalling that across levels $a' < a$)

$$\mathrm{Cov}\left(\frac{\Psi_2(a)}{\sigma(2,a)}, \frac{\Psi_1(a')}{\sigma(1,a')}\right) = \left(aa'\sigma_\omega^2 + \sigma_\Upsilon^2 + (a' - \underline{a}(1))\sigma_{\varepsilon(1)}^2\right)\frac{1}{\sigma(2,a)\sigma(1,a')},$$
$$a > \bar{a}(1);\ \underline{a}(1) < a' \leq \bar{a}(1).$$

This adds no new information given the initial normalization.

Under conditions established in Matzkin (2007) and Heckman and Vytlacil (2007), we can nonparametrically identify the distributions of $\varepsilon(1,a)$ and $\varepsilon(2,a)$ for each $a$ in the appropriate intervals and the technologies at each level subject to the initial normalization. We do not develop this point further because we assume parametric models in making our estimates. The conditions just developed extend in a straightforward way to higher levels, $\ell > 2$. All higher-level parameters are identified up to the initial normalization.

### 5.2.1 Testing the Scale Invariance Assumption

Under scale invariance, we obtain tight restrictions on the coefficients across levels. Relaxing scale invariance adds two new parameters $(\gamma_{0,2}, \gamma_{1,2})$ to Equation (10):

$$\ln K(2,a) = \gamma_{0,2} + \mu_2 + V_2(Q(a)) + \bar{\beta}_2(X)(a - \bar{a}(1)) + \sum_{j=\underline{a}(2)}^{a} \varepsilon(2,j) + \gamma_{1,2}\ln K(1,\bar{a}(1)).$$

Notice that it imposes a proportionality restriction across functions common to $\ln K(2,a)$ and $\ln K(1,a)$. Going across levels,

$$\mathrm{Cov}\left(\frac{\Psi_2(a)}{\sigma(2,a)}, \frac{\Psi_1(a')}{\sigma(1,a')}\right) = \gamma_{1,2}\left\{aa'\sigma_\omega^2 + (a' - \underline{a}(1))\sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2\right\}\frac{1}{\sigma(2,a)\sigma(1,a')},$$

$$a > \bar{a}(1);\ \underline{a}(1) \leq a' < \bar{a}(1).$$

From the previous analysis, the term in braces is identified (up to the previous normalization) at the first level. Thus $\gamma_{1,2}$ is identified, and we can test if $\gamma_{1,2} = 1$. Testing $\gamma_{0,2} = 0$ requires stronger assumptions. We need model intercepts to be invariant, which is difficult to maintain given that $\bar{K}(2)$ is absorbed in any estimated intercept, and we expect that the difficulty levels are increasing in $\ell$. As before, we can estimate $\ln \bar{K}(2)$ up to scale net of intercepts, and we can identify the scale.

# 6 Estimation Results

We use the method of simulated moments to estimate the model for each specific skill $s$. We adjust for clustering in our sample using the paired cluster bootstrap. Details are provided in Appendix P. The moments used in forming the estimates are presented in Table Q1. The model passes goodness of fit tests (see Appendix Q). Appendix Q also plots model predictions vs data for each skill, with and without scale invariance (see Figures Q1, Q7, and Q13 for language, cognition, and fine motor skills, respectively). In general, imposing scale invariance produces worse fits, a point developed further below. We report estimates in the text, relaxing scale invariance. Estimates imposing scale invariance are in Appendix R.

## 6.1 Estimates

The learning component $\delta_\ell E(\eta)$ combines the lesson on $\ell$ (i.e., $\delta_\ell$) taught at age $a$ that is common across all children and $\eta$, the child's ability. The intervention interaction measures (entered as $X$ in $\beta_\ell(X)$) significantly improve child learning for each task. Consistent with the results reported for our analysis of the treatment effects for our new measures of skill, the interaction between the home visitor and the caregiver is the only consistently positive interaction in promoting skills (see Table R4 in Appendix R).[20] The grandmother as the main caregiver often exhibits significantly negative effects, as was the case for our analysis of our new skill measures.[21] Rapid learning (high-ability) children have significantly higher values of the learning component during the intervention for all skills. This finding is consistent across all difficulty levels for all skills (see Figure 9). We also find that higher caregiver education levels significantly increase the language skill when the children are first enrolled in the program (see Table R1). There is substantial learning for children exposed to more educated mothers.

The estimates do not support age invariance across all levels, but an age-invariance assumption cannot be rejected for some levels and some skills. For example, we cannot reject age invariance between levels 8, 9, and 10 for language skill tasks. Age invariance appears to be a valid description of fine motor skills at all levels. We find that the difficulty levels with large $\sigma(\varepsilon, \ell)$ have a large range of passing rates. Passing rates do not always increase monotonically by task order within the same level.

[20] All the estimation results are presented in Appendix R.

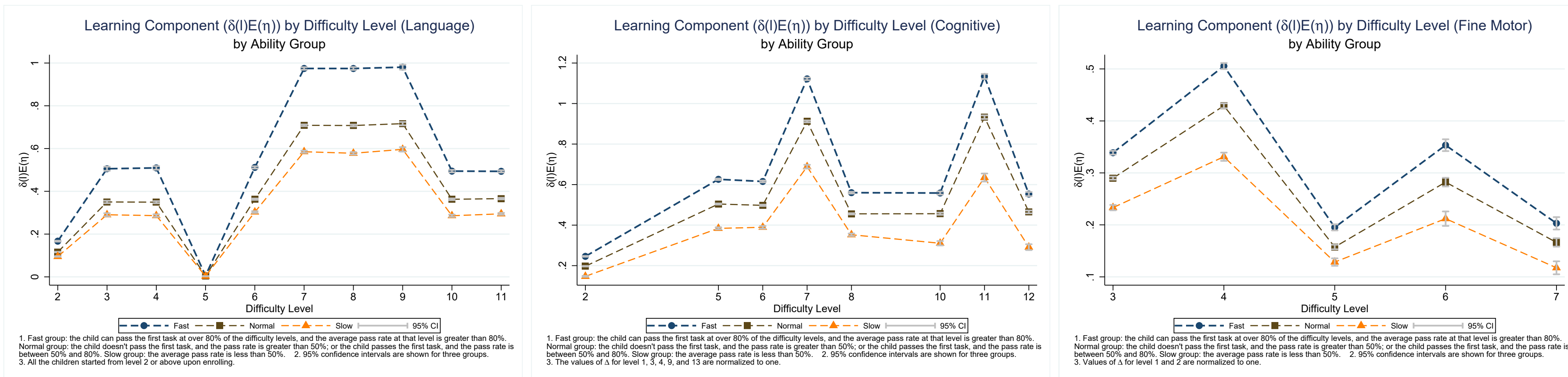
[21] Grandmothers' education is low on average (3 years).

## 6.2 Learning Components and Task Performance

Task passing rates increase within difficulty levels. In Zhou, Heckman, Wang, and Liu (2021), we use exchangeability tests on learning patterns to formally test the existence of learning. In this subsection, we examine how the learning component in our structural model explains child task performance.

Figure 9: Estimates of $\delta(\ell)E(\eta)$ by Ability Group

(a) Language (b) Cognitive (c) Fine Motor




* Intervals are of the form $(j-1, j]$. The parameter for the interval is indexed by the upper value, $j$.

In our setting, the learning component is $\delta_\ell \eta(X)$. The $\delta_\ell$ term captures the curriculum content at each difficulty level, and the $\eta(X)$ term includes the interaction quality, teaching quality, and grandmother quality during the intervention. Therefore, we focus on how the $\eta(X)$ term affects child task performance. Figure 10a shows the mean of $\eta(X)$ for each task. We identify it using $\beta_\ell$ and normalizing $\delta(1) = 1$. There is an increasing pattern of $E(\eta)$ within difficulty levels. In Figure 10b, we break down the estimated $E(\eta)$ values by ability group. Children in the normal group contribute the most growth to the learning component. Children in the fast group master the task quickly, usually on the first try. Thus, there is little subsequent growth in learning for them when they are instructed on the same task multiple times. For children in the normal group, performance improves as they learn the task multiple times. This is consistent with our estimates showing that the learning component $E(\eta)$ increases within a difficulty level, especially strongly for children in the normal group. This finding is also found for cognitive tasks. For fine motor tasks, there is a similar pattern for tasks greater than 4, although learning is not substantial at any level.

Figure 10a: Learning Component $E(\eta(X))$ of Language Tasks by Level

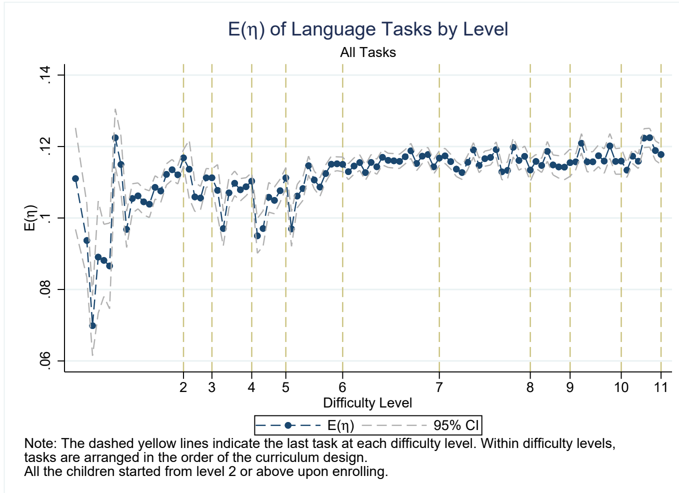



Figure 10b: Learning Component $E(\eta(X))$ of Language Tasks by Level and Ability Group

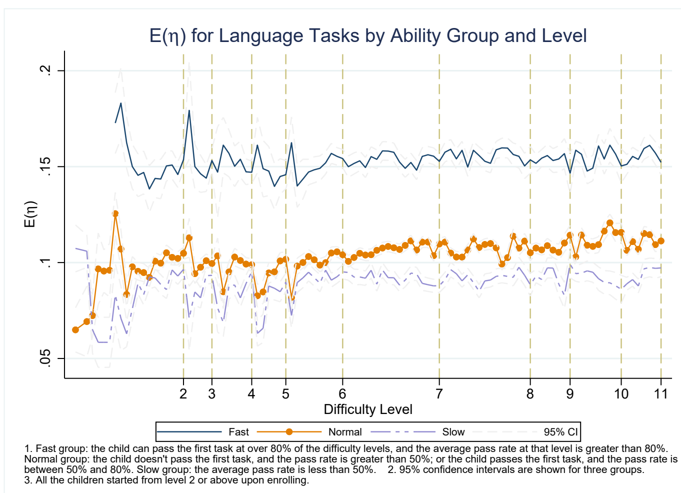
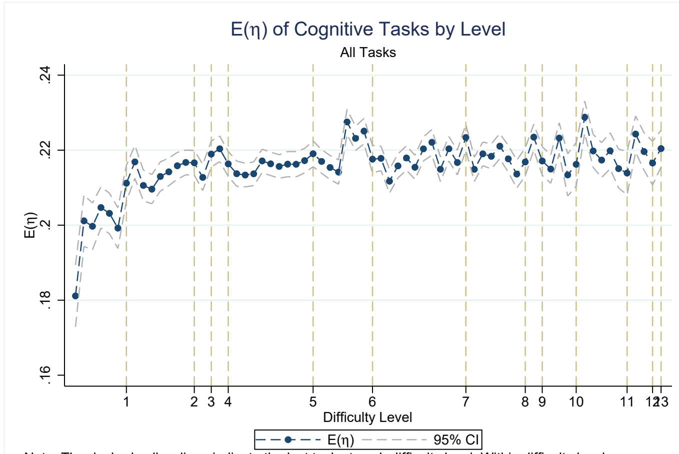
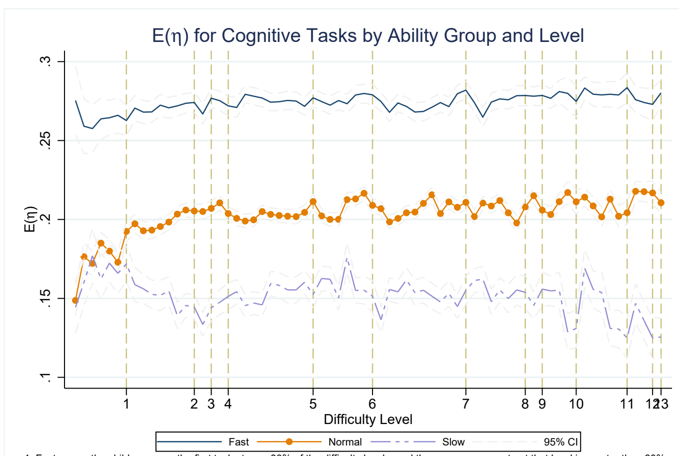


Figure 11a: Learning Component $E(\eta(X))$ of Cognitive Tasks by Level




Note: The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the curriculum design.

Figure 11b: Learning Component $E(\eta(X))$ of Cognitive Tasks by Level and Ability Group




1. Fast group: the child can pass the first task at over 80% of the difficulty levels, and the average pass rate at that level is greater than 80%. Normal group: the child doesn't pass the first task, and the pass rate is greater than 50%; or the child passes the first task, and the pass rate is between 50% and 80%. Slow group: the average pass rate is less than 50%. 2. 95% confidence intervals are shown for three groups.

Figure 12a: Learning Component $E(\eta(X))$ of Fine Motor Tasks by Level

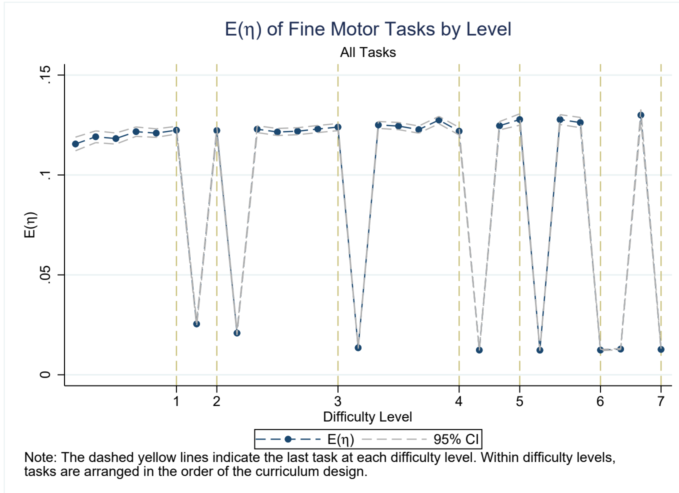
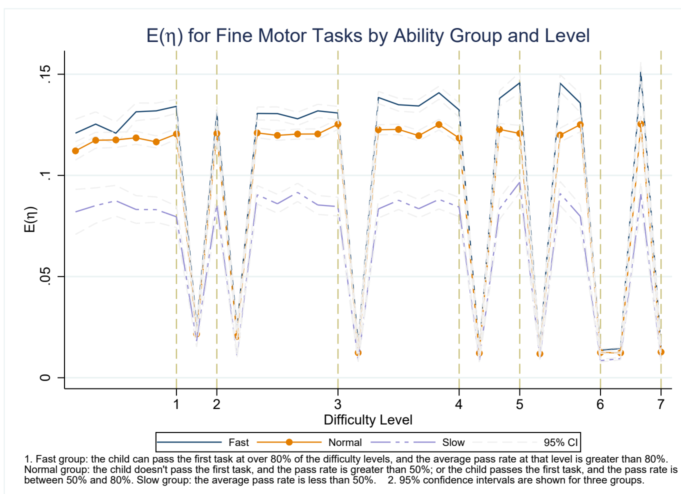



Figure 12b: Learning Component $E(\eta(X))$ of Fine Motor Tasks by Level and Ability Group



## 6.3 Language Skills

Figure 13(a) displays estimates of the estimated minimum skill requirement for each level. This is defined relative to $\bar{K}(1)$, assuming no shift in model intercepts for each skill across levels. As expected, the skill required to pass tasks is monotonically increasing across difficulty levels. The variance of shocks at each level displays different patterns, reflecting differential ability. Figure 13(b) presents the estimates of variances, and the variances at levels 6, 8, and 11 are larger than the variances at other levels. We plot the task passing rates at these three levels in Figure 14, and we find that the large variances are associated with a larger range of passing rates. Passing rates do not monotonically increase by task order within the same level (see Figure 14). Level-specific shocks intrude to alter the monotonicity delivered by the deterministic component of the model and to reflect the lack of fit of the model to the data (see Figure Q1(b)).

Figure 13: Language Skill

(a) $\bar{K}(\ell)$ (b) $\sigma^2(\varepsilon, \ell)$

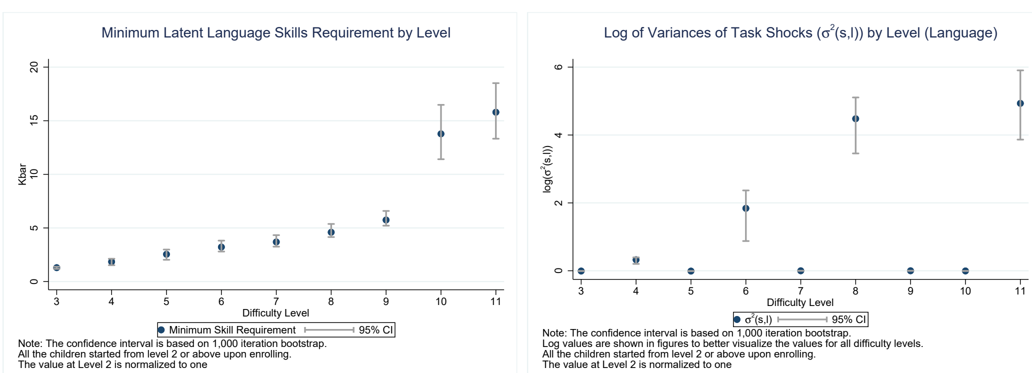
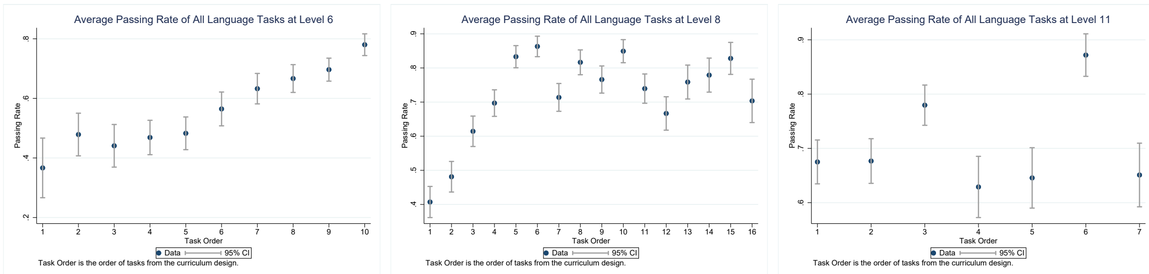


Figure 14: Average Passing Rate of Language Tasks: $p(s, \ell)$ (Raw Data)

(a) Level 6 (b) Level 8 (c) Level 11




## 6.4 Cognitive Skills

The pattern for the estimated parameters for cognitive skills is similar to that for language skills. For certain difficulty levels, passing rates are not monotone within levels.

Figure 15: Cognitive Skill

(a) $\bar{K}(\ell)$ (b) $\sigma^2(\varepsilon, \ell)$

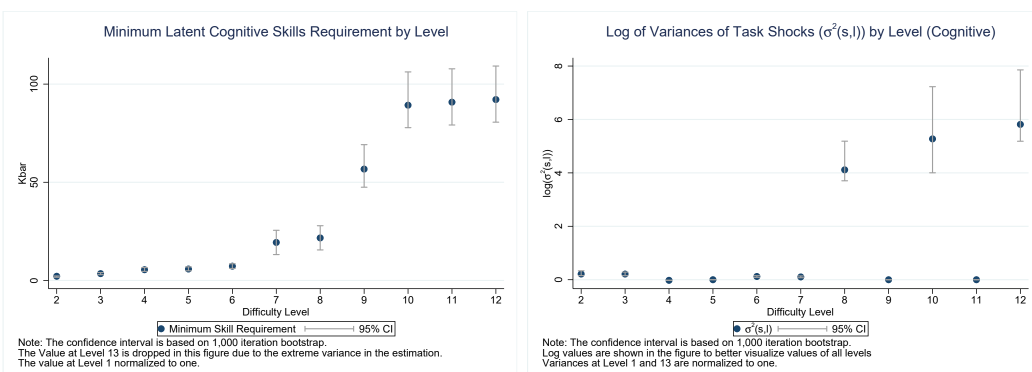
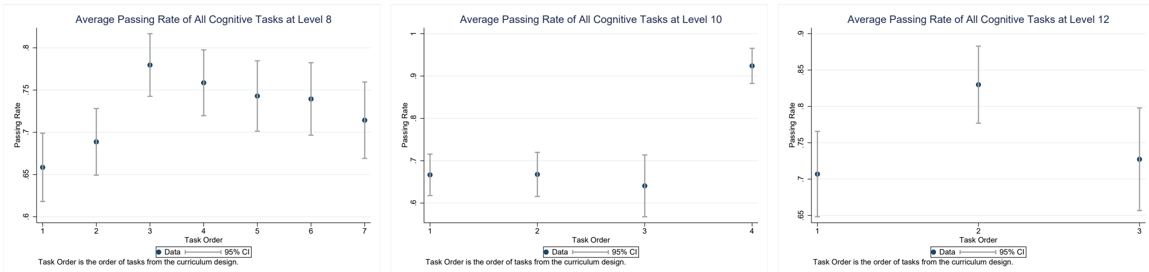


Figure 16: Average Passing Rate of Cognitive Tasks: $p(s, \ell)$ (Raw Data)

(a) Level 8 (b) Level 10 (c) Level 12




## 6.5 Fine Motor Skills

A similar pattern arises for fine motor skill.

Figure 17: Fine Motor Skill

(a) $\bar{K}(\ell)$ (b) $\sigma^2(\varepsilon, \ell)$

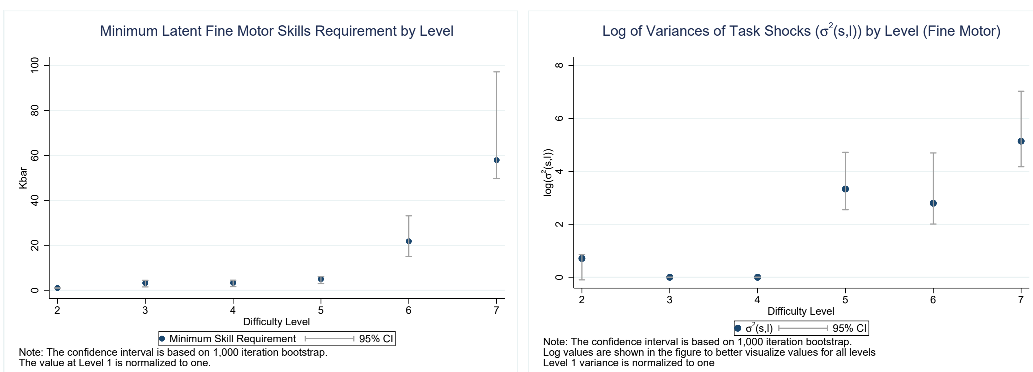
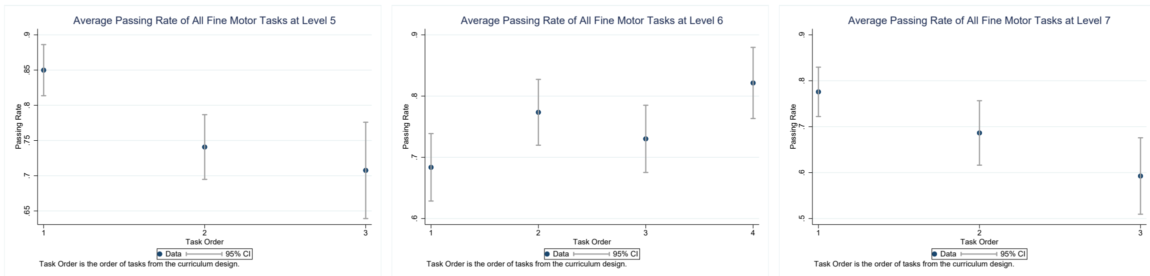


Figure 18: Average Passing Rate of Fine Motor Tasks: $p(s, \ell)$ (Raw Data)

(a) Level 5 (b) Level 6 (c) Level 7




## 6.6 Testing the Scale Invariance Assumption

Under scale invariance, $\gamma_{1,\ell} = 1$. Note that $\gamma_{1,\ell} = 1$ means a uniform scale across $\ell$ and $\ell - 1$. Figure 19a shows estimates of $\gamma_{1,\ell}$ for each level for a model estimated without imposing that restriction. Table 8 shows the $\chi^2$ test results for each level and skill. Our estimates partially support scale invariance. For some levels, scale invariance cannot be rejected. For example, we cannot reject scale invariance for language skills between levels 8-11 (i.e., 8-9, 9-10, and 10-11). It is decisively rejected at levels 4-6. Table 9 lists the task content for difficulty level 8-11; it shows that the task content is very similar across these different levels.

We also test for scale invariance for cognitive and fine motor skill tasks. We reject scale invariance for all the levels of the cognitive skill tasks. However, we find evidence in support of skill invariance for fine motor skill tasks, which mainly focus on drawing skills.

Figure 19: Tests of the Null Hypothesis of Scale Invariance

(a) Language Skill (b) Cognitive Skill (c) Fine Motor Skill

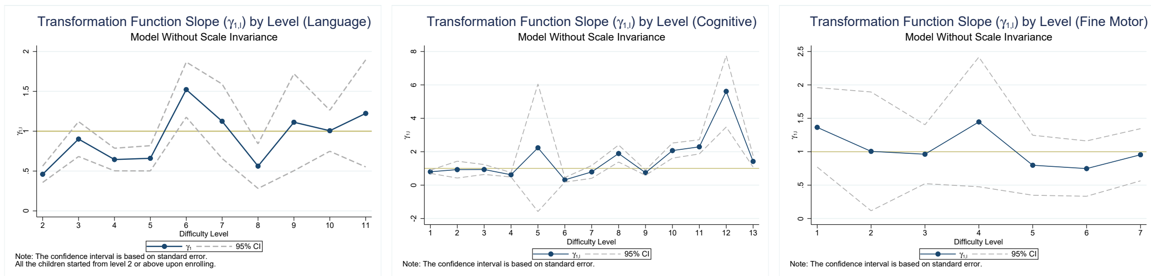



Table 8: Scale Invariance Hypothesis Tests by Levels

| | Language | | | Cognitive | | | Fine Motor | | |
|---|---|---|---|---|---|---|---|---|---|
| | Slope($\gamma_{1,\ell}$) | $\chi^2(1)$ | $p$-value | Slope($\gamma_{1,\ell}$) | $\chi^2(1)$ | $p$-value | Slope($\gamma_{1,\ell}$) | $\chi^2(1)$ | $p$-value |
| Level 1 | | | | 0.800 | 3.946 | 0.047 | 1.365 | 1.414 | 0.234 |
| Level 2 | 0.46 | 57.846 | 0.000 | 0.929 | 0.012 | 0.914 | 1.005 | 0.000 | 0.992 |
| Level 3 | 0.901 | 0.546 | 0.460 | 0.936 | 0.010 | 0.922 | 0.963 | 0.022 | 0.883 |
| Level 4 | 0.645 | 20.193 | 0.000 | 0.621 | 0.142 | 0.707 | 1.446 | 0.774 | 0.379 |
| Level 5 | 0.66 | 9.382 | 0.002 | 2.235 | 3.899 | 0.048 | 0.798 | 0.720 | 0.396 |
| Level 6 | 1.522 | 5.063 | 0.024 | 0.317 | 17.482 | 0.000 | 0.748 | 1.277 | 0.258 |
| Level 7 | 1.125 | 0.182 | 0.670 | 0.791 | 0.362 | 0.547 | 0.955 | 0.034 | 0.853 |
| Level 8 | 0.562 | 8.195 | 0.004 | 1.893 | 4.237 | 0.040 | | | |
| Level 9 | 1.113 | 0.113 | 0.737 | 0.744 | 3.432 | 0.064 | | | |
| Level 10 | 1.006 | 0.001 | 0.970 | 2.068 | 12.211 | 0.000 | | | |
| Level 11 | 1.223 | 0.375 | 0.540 | 2.292 | 10.927 | 0.001 | | | |
| Level 12 | | | | 5.614 | 14.351 | 0.000 | | | |
| Level 13 | | | | 1.420 | 4.333 | 0.037 | | | |
| Total | | 101.897 | 0.000 | | 75.344 | 0.000 | | 4.241 | 0.752 |

1. For each level we test the null hypothesis that $\gamma_{1,\ell}$=1 .
2. The column of $p$-value reports the probability of not rejecting the null hypothesis.
3. The row "Total" tests whether the scale invariance assumption is valid across all the levels.
4. Our data for language tasks starts from level 2.

The evidence for skill invariance across levels 8-9, 9-10, and 10-11 makes sense given the similarity of the tasks at those levels.

Table 9: Difficulty Level List for Language (Learn words) Tasks

| | |
|---|---|
| Level 8 | The child points to the pictures which are being named, names one or more pictures, and mimics the sound of the objects. |
| Level 9 | The child points to the pictures which are being named, names two or more pictures, makes the sound of the objects. |
| Level 10 | The child points at 7 or more than 7 pictures and talks 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. |

Our evidence on scale invariance is based on a parametric normal specification. This limits the generality of our findings. As previously noted, it is possible to estimate a nonparametric version of the model. This is a task left for the future.

# 7 Conclusion

This paper uses novel experimental data on a widely-emulated home visiting program implemented in rural China. We study its mechanisms for improving child skills. We investigate the impacts of different types of interactions on child achievement measures: interactions between the home visitor and the caregiver, interactions between the home visitor and the child, the quality of the teacher, and the frequency of the caregiver playing with the child after the class. We find that high-quality interactions between the home visitor and the caregiver significantly improve child skill development in multiple dimensions, but the program's other features are not generally effective.

We develop and estimate new skill learning and ability measures that extend the traditional measure (% of correct answers) in two dimensions: time to mastery—how fast the child achieves task mastery—and backsliding, which examines how stable the child's performance is after first mastery. We propose and implement tests for dynamic complementarity that do not rely on arbitrary measures of skill. We test and reject a unitary skill model and in fact find as many as 5 dimensions of skill, including multiple dimensions within nominally identical skill categories.

We develop and estimate a dynamic learning model consistent with our evidence on treatment effects for various measures of learning. The model captures the patterns of learning in our data and explains how skills evolve at weekly levels. We measure the growth in knowledge across ability levels. Because lessons are the same for all children of the same age, normal-ability children experience more learning patterns. $E(\eta)$ is higher for high-ability children but does not improve within the same difficulty level because they generally master the task after their first lesson. Going forward, in designing the program, adaptive lessons that accelerate for high-ability children will promote greater learning.

We formally test whether age invariance holds across each level. We find evidence supporting age invariance for certain skills at certain difficulty levels, but reject the scale invariance assumption as a global characterization, except for fine motor skills. This finding calls into question standard practice that relies on age-invariant measures for analyzing child development and the value added of teachers and schools.

# References


Agostinelli, F. and M. Wiswall (2021). Estimating the technology of children's skill formation. Revision requested, Journal of Political Economy.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, Chapter 1, pp. 1–181. Amsterdam: Elsevier B. V.

Andrew, A., O. Attanasio, B. Augsburg, M. Day, S. Grantham-McGregor, C. Meghir, F. Mehrin, S. Pahwa, and M. Rubio-Codina (2019). 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*.

Bai, Y. (2019). Optimality of matched-pair designs in randomized controlled trials. Unpublished manuscript, University of Chicago.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016). What do grades and achievement tests measure? Submitted to *Proceedings of the National Academy of Sciences*.

Britto, P. R., P. L. Engle, and C. M. Super (Eds.) (2013). *Handbook of early childhood development research and its impact on global policy.* Oxford University Press. Sponsored by UNICEF and SRCD.

Bronfenbrenner, U. (Ed.) (2005). *Making Human Beings Human: Bioecological Perspectives on Human Development.* SAGE Publications.

Cunha, F. and J. J. Heckman (2008, Fall). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Cunha, F., E. Nielsen, and B. Williams (2021). The econometrics of early childhood human capital and investments. *Annual Review of Economics 13*(1), 487–513.

Del Boca, D., C. J. Flinn, and M. Wiswall (2014). Household choices and child development. *Review of Economic Studies 81*(1), 137–185.

Doepke, M. and F. Zilibotti (2019). *Love, Money, and Parenting: How Economics Explains the Way we Raise our Kids*. Princeton University Press.

Elango, S., , J. L. García, J. J. Heckman, and A. Hojman (2016). Early childhood education. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States*, Volume 2, Chapter 4, pp. 235–297. Chicago: University of Chicago Press.

Engle, P. L., L. C. H. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. Cabral de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, and S. Iltus (2011, October). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet 378*(9799), 1339–1353.

Freyberger, J. (2021). Normalizations and misspecification in skill formation models. arXiv preprint arXiv:2104.00473.

Gertler, P., J. J. Heckman, R. Pinto, S. M. Chang-Lopez, S. Grantham-McGregor, C. Vermeersch, S. Walker, and A. S. Wright (2021). Effect of the jamaica early childhood stimulation intervention on labor market outcomes at age 31. NBER 29292. Under review.

Gertler, P., J. J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. Chang, and S. M. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science 344*(6187), 998–1001.

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

Heckman, J. J. (1978, September). Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales d l'INSEE 30–31*, 227–269. Special Issue.

Heckman, J. J. (1981). Statistical models for discrete panel data. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 114–178. Cambridge, MA: MIT Press.

Heckman, J. J. and S. Mosso (2014, August). The economics of human development and social mobility. *Annual Review of Economics 6*(1), 689–733.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 70, pp. 4779–4874. Amsterdam: Elsevier B. V.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Technical report, Organisation for Economic Co-operation and Development, Paris. Available at https://www.oecd.org/edu/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non-Cognitive-Skills-to-Promote-Lifetime-Success.pdf.

Kim, J. H. (2019). Parenting skill and child development. Unpublished.

Lord, F. M. and M. R. Novick (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.

Matzkin, R. L. (1992, March). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica 60*(2), 239–270.

Matzkin, R. L. (2007). Nonparametric identification. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B. Amsterdam: Elsevier.

OECD (2021). *Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills*. OECD Publishing.

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

Rutherford, R. S. G. (1955, July). Income distributions: A new model. *Econometrica 23*(3), 277–294.

Thelen, E. (2005, 04). Dynamic systems theory and the complexity of change. *Psychoanalytic Dialogues 15*, 255–283.

Todd, P. E. and K. I. Wolpin (2007, Winter). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital 1*(1), 91–136.

Uzgiris, I. C. and J. Hunt (1975). Assessment in infancy: Ordinal scales of psychological development. *University of Illinois Press*.

van der Linden, W. J. (2016). *Handbook of Item Response Theory: Volume 1: Models.* CRC Press.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press.

Zhou, J., J. Heckman, B. Liu, and M. Lu (2021). Treatment effects and the measurement of skills in an influential home visiting program. Working Paper.

Zhou, J., J. Heckman, F. Wang, and B. Liu (2021). Tests of early childhood skill learning patterns. Revision requested, *Journal of Community Psychology.*