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
December 29, 2022

---

[*] 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), the Bonn IZA-Briq Life-Cycle Workshop (August 19, 2021), and at Rice University (November, 2021). We have benefited from the comments of the editor, John List, and four insightful referees, as well as those of our discussant Lance Lochner at a September 2019 Becker Friedman Institute conference in Chicago, as well as those of Flavio Cunha, and Pat Kyllonen. Fuyao Wang, Haihan Tian, and Alejandra Campos contributed highly competent and insightful research assistance and commentary.

Abstract

This paper uses novel experimental data from a widely-emulated early childhood home-visiting program implemented in China with high-frequency measurements to investigate the dynamics of skill formation. We show that home visiting interventions promoting child development promote quality interactions between home visitors and caregivers. We report non-parametric evidence supporting dynamic complementarity-that early investment makes later investment more productive-that does not rely on arbitrary measures of skills. Based on this evidence, we formulate and estimate a dynamic stochastic learning model for multiple skills and quantify the sources of early life learning. Our model unites and extends two basic and widely-used psychometric models of learning and measurement: the IRT model and the BKT model. Using our model, we test the widely held assumption of the existence of constant unit measures of skill ("human capital") that are comparable over levels of skills and across ages for language and cognitive latent skills. We find evidence supporting it for certain levels of skill but not for all. Our stochastic growth model explains the frequently observed phenomena of "fadeout" of interventions.

JEL Codes: I3, J1, C5, D2, O12, C9

Keywords: child development, measures of skills, scaffolding, targeting, experiment, IRT, BKT

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

This paper analyzes a low-cost home visiting program in China with unique high-frequency (weekly) data on skill development that is based on a widely-emulated program 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, 2022). The study of the effectiveness of home visiting programs isolates a component of successful omnibus programs that include this feature (see Zhou et al., 2022; García and Heckman, 2022).

We investigate the mechanisms producing growth of knowledge on multiple skills in the early years. Using nonparametric methods, we find evidence consistent with a crucial property of learning: dynamic complementarity–acquired skills foster the growth of later skills. We develop and estimate a micro-dynamic model of reinforcement learning to characterize the dynamics of skill formation during early childhood.

The *technology of skill formation* (Cunha and Heckman, 2007) characterizes the growth of child skills at age (stage) $\boldsymbol{a} : \boldsymbol{K}(\boldsymbol{a})$. It is a function of a vector of investments $\boldsymbol{I}(\boldsymbol{a})$ (including home visits, parenting, interactions with the child, school-based interventions, center care, school stimulation, etc.) and environments $\boldsymbol{G}(\boldsymbol{a})$ (including neighborhoods, parental education, and public goods):

$$\overbrace{\boldsymbol{K}(\boldsymbol{a}+\boldsymbol{1})}^{\substack{\text{skills at}\\ \boldsymbol{a+1}}} = \boldsymbol{f}^{(\boldsymbol{a})}\Big(\underbrace{\boldsymbol{K}(\boldsymbol{a})}_{\substack{\text{Skills at}\\ \boldsymbol{a}}}, \overbrace{\boldsymbol{I}(\boldsymbol{a}+\boldsymbol{1})}^{\text{Investment}}, \underbrace{\boldsymbol{G}(\boldsymbol{a}+\boldsymbol{1})}_{\substack{\text{Environmental}\\ \text{Variables}}}\Big). \tag{1}$$

For simplicity we assume this age- or stage-dependent function is twice continuously differentiable. Key properties of $f^{(a)}$ featured in the literature are self productivity ($\frac{\partial \boldsymbol{K(a+1)}}{\partial \boldsymbol{K(a)}} > 0$), the productivity of investments and beneficial environments ($\frac{\partial \boldsymbol{K(a+1)}}{\partial \boldsymbol{I(a+1)}} > 0$, $\frac{\partial \boldsymbol{K(a+1)}}{\partial \boldsymbol{G(a+1)}} > 0$), and critical and sensitive periods of development ($\boldsymbol{f_2^{(a)}} > \boldsymbol{f_2^{(a')}}$, $a \neq a'$, evaluated at common values, where $\boldsymbol{f_2^{(a)}}$ is $\frac{\partial \boldsymbol{K(a+1)}}{\partial \boldsymbol{I(a+1)}}$). Static complementarity ($\frac{\partial^2 \boldsymbol{K(a+1)}}{\partial \boldsymbol{K(a)} \partial \boldsymbol{I(a+1)}} > 0$) is often found in empirical studies of child development. Investment is more productive the higher the stock of skills; i.e., "skill begets skill." Dynamic complementarity ($\frac{\partial^2 \boldsymbol{K(a+j+1)}}{\partial \boldsymbol{I(a)} \partial \boldsymbol{I'(a+j)}} \geq 0$, for $j > 1$) is a central proposition in the literature. It asserts that investment at earlier life cycle stages makes later investments more productive. It implies that remediation of skill deficits is more costly (requires more investment) at later stages of the life cycle than direct investment at early ages (Heckman and Mosso, 2014).

There are three big questions in this literature. (1) What is $\boldsymbol{I(a)}$ and how to measure it?[1] (2) What are the micro-mechanisms underlying the technology? Child psychologists emphasize that warm and supportive parent/caregiver-child interactions—*"scaffolding"* (Vygotsky, 1978)—are major determinants of child development. (3) How should we measure skills and their growth?

This paper focuses on the mechanisms underlying technology (1) using high frequency (weekly) data on the growth of skills in the treatment group of the China REACH home visiting program. The paper by Zhou, Heckman, Liu, and Lu (2022), examines treatment effects of the program studied at endline and midline and presents and applies methods for correcting for item difficulty. Zhou, Heckman, Liu, Lu,

---

[1] Many different definitions are used. 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 (2022); Andrew et al. (2020); Doepke and Zilibotti (2019).

Chang, and Grantham-McGregor (2022) consider the issue of external validity and compliance with the SANS conditions of List (2020). A third paper focuses on issues of measurement (Heckman and Zhou, 2022a).

In the literature, test scores based on passing rates on assessments of cognitive, socioemotional, and other skills are widely used.[2] 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 do not suffer from this problem, but at the same time, do not measure levels of attained skill. Value-added measures of school, teacher and student quality assume that constant unit measures are available to make meaningful comparisons.[3]

As a byproduct of our dynamic model, we propose and implement model-based tests of *invariance* of latent skills, a crucial assumption maintained in the value-added and human capital literatures and specifically in previous research on skill formation. It maintains the existence of constant-unit latent skills (human capital) over all levels of the same skill. This literature also assumes the existence of constant-unit *measures* of latent skills, which may or may not exist even if constant unit latent skill scales exist.[4] One approach to the problem of defining and measuring

---

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

[3] Cunha et al. (2010, 2021); García and Heckman (2022); Agostinelli and Wiswall (2022); Freyberger (2021).

[4] For example, Todd and Wolpin (2007) and others use words spoken by age as a measurement of the invariant latent skill. 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 percent changes comparable at different ages? What is the appropriate metric? Are there common scales of knowledge? Is there a single scale to measure the growth of knowledge over time? For all skills? For any particular skill? An assumption of common scales of measurement ignores the finding that multiple skills emerge as a child matures. 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).

scale anchors test scores in meaningful outcomes (e.g., earnings, crime). However, objective behavioral anchors at early ages are difficult to find.[5] The recent literature demonstrates the empirical importance of these issues. Freyberger (2021) shows the dramatic consequences of different scalings of skill measures for estimates of the technology of Equation (1).

The current paper addresses the first two questions and the first aspect of the third problem (Existence of invariant latent scales). In Heckman and Zhou (2022a), we address the second aspect: the issue of existence of invariant *measures* of skills.

The paper proceeds as follows. We report empirical evidence on the key features of the dynamics of the learning process. We examine key mechanisms of home visiting interventions that improve child skill development. We evaluate the impacts on child development of the interactions between home visitors and caregivers and the impact of home visitors' teaching quality. We present evidence consistent with dynamic complementarity without imposing constant-unit invariance assumptions.

We develop and estimate a new stochastic micro-dynamic model of skill formation that formalizes mechanisms proposed in developmental psychology and explains uneven growth of test scores over levels and fadeout.[6] We investigate the growth of skills at more granular levels than previous analyses.

We report the following findings. (1) A key mechanism fostering growth of child skills is quality interaction between home visitors and caregivers. (2) We present evidence consistent with dynamic complementarity using nonparametric methods. (3) Based on this evidence, we develop and estimate a dynamic reinforcement learning

---

[5] For a recent discussion of these problems, see Cawley et al. (1998) and Cunha et al. (2021).

[6] See, e.g., Bronfenbrenner (2005) and Thelen (2005).

model. It unites and extends two highly-influential models of psychometrics: the IRT (Item Response Theory) model and the BKT (Bayesian Knowledge Tracing) model.[7] We add investment and stochastic growth to these frameworks. We find evidence supporting the assumption of invariance of latent skills across levels for certain skills at specific skill levels but not globally. This complements our analysis on invariant *measures* in Heckman and Zhou (2022a).

The paper unfolds in the following way. Section 2 describes the background of the program we analyze and its curriculum.[8] Section 3 presents our evidence on learning patterns. Section 4 discusses the impacts of different interactions on learning. Section 5 presents nonparametric evidence consistent with dynamic complementarity. Section 6 develops a latent Markov process micro-dynamic learning model. Section 7 presents estimates and interpretations. Section 8 concludes.

2 China REACH

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. Substantial positive effects are found for the program through age 34 (i.e., Gertler et al., 2022, 2014). Its success has spawned replications around the world, e.g., in Bangladesh, China, Colombia, India, Peru (see, e.g., Grantham-McGregor and Smith, 2016).

---

[7] See van der Linden (2016).

[8] Zhou et al. (2022) describe it in much greater detail.

The program we analyze, *China REACH*, extends and applies the Jamaican protocols. 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. This intervention is not focused on stunted children.

China REACH is a paired-match RCT that minimizes mean square errors of estimates (Bai et al., 2021; Bai, 2022). A non-bipartite Mahalanobis matching method[9] 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. More details of the design of the experiment and the balance test between treatment and control groups can be found in Zhou, Heckman, Liu, and Lu (2022).

The intervention focuses on improving multi-dimensional skill development through a home visit delivery model. Trained home visitors who are roughly at the level of education of the mothers of the children studied visit each treated household weekly and provide one hour of caregiving guidance.

Zhou, Heckman, Liu, and Lu (2022) evaluate the treatment effects of the intervention and find that the intervention significantly improves skill development (e.g., language and cognitive, fine motor, and social-emotional skills). To interpret treatment effects, they use item responses on inventories of skill to estimate individual latent skills. They decompose the source of treatment effects and find that enhancement in latent skills explains most of the conventional treatment effects. Zhou, Heckman, Liu, Lu, Chang, and Grantham-McGregor (2022) show that the skill profiles

---

[9] See Lu et al. (2011).

for the growth of skills are similar to those of the original Jamaica Reach program, suggestions some generalizability.

2.1 Program Protocols

The program 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 literature on child development. Central to our identification strategy is the assumption that these profiles describe valid hierarchies (levels) of knowledge and that the knowledge content is the same within each level.[10] Child skills are assessed weekly. There are monthly assessments of the quality of home visits recorded by supervisors.

There are 13 difficulty levels for cognitive skills. Table 1 gives the tasks for cognitive skills taught at specific levels and Figure 1 presents the timing of the lessons by age. The tasks start with 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 described in Table 2. All tasks at that level are virtually identical in task difficulty and relate to the activ-

---

[10] The difficulty levels are ordered based on the average children's performance. (see Palmer, 1971)

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

ity of looking at pictures or objects and vocalizing. Appendix A gives comparable information for the other skills.

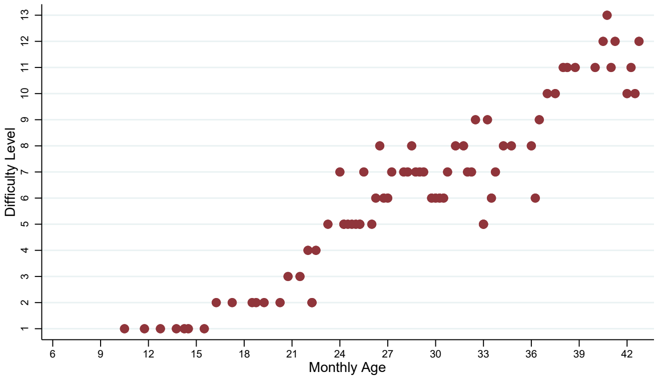
3 Empirical Evidence on Learning

This section documents the observed learning patterns of the study.

3.1 High Frequency Data on Learning

Our data on weekly skill growth enable us to move beyond traditional aggregate measures such as the percent of items passed over a diverse range of tasks to exam-

Figure 1: The Timing of Cognitive Skill (Understand 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.

ine task by task skill growth and the factors that influence it. To understand the structure of the data analyzed, we introduce some helpful notation.

Let $\mathcal{S}$ be the set of skills taught. Let $\ell(s, a)$ be the level of skill s taught at age a. Mastery of skill s at level ℓ at age a is characterized by:

$$D(s, \ell, a) = \begin{cases} 1 & K(s, \ell, a) \geq \bar{K}(s, \ell) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

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 ℓ. This characterization is consistent with the classical IRT model (Lord and Novick, 1968). Let $\underline{a}(s, \ell)$ be the first age at which skill s is taught at level ℓ, and let $\bar{a}(s, \ell)$ be the last age at which it is taught at level ℓ. 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 ℓ) starting at age $\underline{a}(s, \ell)$. For

Table 2: Cognitive Skill Task Content: Look at the Pictures and Vocalize (Level 1)

Difficulty Level	Difficulty Level Aim	Month	Week	Learning Materials	Task Aim and Content
Level 1	Look at the pictures and vocalize	10	2	Picture book A	Look at the pictures and vocalize: baby makes sound when looking at the pictures
Level 1	Look at the pictures and vocalize	11	3	Picture book B	Look at the pictures and vocalize: baby looks at the pictures and vocalize
Level 1	Look at the pictures and vocalize	12	3	Picture book A	Look at the pictures and vocalize: baby makes sound when looking at the pictures
Level 1	Look at the pictures and vocalize	13	3	Picture book B	Look at the pictures and vocalize: baby looks at the pictures and vocalize
Level 1	Look at the pictures and vocalize	14	1	Picture book A	Look at the pictures and vocalize: baby makes sound when looking at the pictures
Level 1	Look at the pictures and vocalize	14	2	Baby doll	Look at the pictures and vocalize: baby makes sound when holding a baby doll
Level 1	Look at the pictures and vocalize	15	2	Picture book B	Look at the pictures and vocalize: The child pronounces while looking at the pictures

level ℓ of skill s, we collect the indicators of knowledge in a spell, $\left\{D(s,\ell,a)\right\}_{\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}$.

3.2 Characterizing Learning

In a stationary environment with age-invariant individual heterogeneity and with no learning or growth of knowledge at level ℓ and skill s, 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 ℓ).[11] With learning, sequences are back-loaded i.e. for $j > 0$, $\Pr(D(s,\ell,a+j) \geq D(s,\ell,a)) \geq \Pr(D(s,\ell,a+j) \leq D(s,\ell,a))$.

Zhou, Heckman, Wang, and Liu (2022) test exchangeability on weekly data and reject that hypothesis, indicating learning. Learning is found even after controlling for maturation and exposure effects that might boost skills in the absence of any intervention (see Appendix B).

Figure 2 characterizes the growth of knowledge in language, cognitive, and fine motor skills.[12] Average passing rates within each difficulty level for language and cog-

---

[11] See Heckman (1978, 1981).

[12] The program has no measured effect on gross motor skills.

nitive tasks increase with age, a pattern consistent with learning. When individuals transition to higher difficulty levels, initial passing rates decline. Subsequent passing rates increase as learning ensues. The dynamic model presented in Section 6 captures this phenomenon. At most levels of fine motor skills, there is—at best—modest learning.[13] Access to detailed weekly data enables us to determine at what stages learning occurs.

3.3 Measures of Learning and Knowledge

Heckman and Zhou (2022a) compare the traditional measure of learning: the proportion of correct answers on a broad range of tasks with two alternative measures of learning and learning speed: time to first mastery and backsliding.

The passing rate on skill s at level ℓ is:

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

The overall passing rate is:

$$p(s) = \sum_{\ell=1}^{L_s} p(s,\ell) w(s,\ell) \tag{4}$$

where L_s is the highest level of skill s and

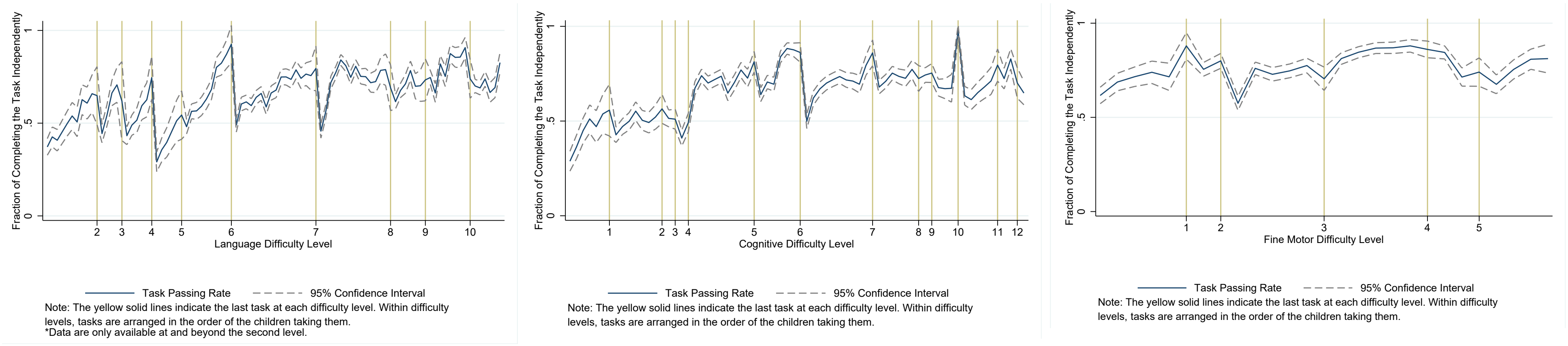
$$w(s,\ell) = \frac{\bar{a}(s,\ell) - \underline{a}(s,\ell) + 1}{\sum_{\ell=1}^{L_s} (\bar{a}(s,\ell) - \underline{a}(s,\ell)) + 1}. \tag{5}$$

This measure weights passing rates at different difficulty levels by the number of items on it tested.

---

[13] 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 2: Average Task Passing Rate by Order and Level



(a) Language

Note: The yellow solid lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.
*Data are only available at and beyond the second level.

(b) Cognitive

Note: The yellow solid lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.

(c) Fine Motor

Note: The yellow solid lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.

There are other plausible measures of knowledge and learning. For consecutive learning spells with all participants entering each level at the first lesson, the ***Time to first mastery*** is $d(s,\ell) = \hat{a}(s,\ell) - \underline{a}(s,\ell)$, where for each s and ℓ, $\hat{a}(s,\ell) = \min_a\{D(s,\ell,a)=1\}_{a=\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}$. This is often used as a measure of intelligence (van der Linden, 2016). Another possible measure is ***Instability*** at level ℓ for skill s as: $\frac{\#\{D(s,\ell,a)=0,a>\hat{a}(s,\ell),a\leq\bar{a}(s,\ell)\}}{\#\{a>\hat{a}(s,\ell),a\leq\bar{a}(s,\ell)\}}\mathbf{1}(\#\{a>\hat{a}(s,\ell), a\leq\bar{a}(s,\ell)\}>0)$.
This captures retention of knowledge.

Heckman and Zhou (2022a) show that in the China REACH data, these measures are correlated in the expected directions. However, the different measures are far from perfectly so, suggesting that they capture different aspects of knowledge.[14] They report that 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 in economics and in the psychology of "g" that claims one universal skill predicts performance on all tasks–is grossly inaccurate.

4 Impacts of Interventions on the Growth of Skills

We now analyze how interventions improve a child's skill development by examining the effects that different interactions have on child learning across difficulty levels. During the intervention, supervisors record assessments of home visitor, caregiver, and child interaction activities at least once per month, making it possible to examine their impacts on skill development. Using these measures, we can evaluate the quality

---

[14] 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 errors in measures of achievement.

of interaction between home visitors and caregivers and between home visitors and children and their impacts. Trained program supervisors evaluate the quality of home visits 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 C describes the interaction data and the factors that summarize it.

Table 3 reports the impact of program interactions on time to the first mastery of achieving cognitive tasks. It shows a recurrent pattern. The interaction between the home visitor and the caregiver is measured at each skill level. We form an average over all visits. It is the only consistently statistically significant pattern across all difficulty levels.[15] Note that age (maturation) effects are statistically important. Children acquire skills with age and experience. Having a grandmother as the caregiver retards learning speed.

In general, the estimated impacts of interactions between home visitors and caregivers on improving children's skills are positive and statistically significant. Estimated impacts of interactions between home visitors and children are generally not significant, nor is the teaching ability of the visitor. To control for any endogeneity that biases home visitor's interactions with children's latent skills by skill level, we measure interaction outcomes for the same home visitor with the children living in different spatially separated villages to construct instruments for the quality of home visitor interaction.[16] When we instrument for home visitor interactions, we find

---

[15] Note that a negative coefficient for a mastery regression means a quicker mastery of the skill.

[16] The instrumental variables include mean, max, and min of other village interaction measures through the same visitor. Details are presented in Appendix D.

Table 3: The Effects of Interactions on the Time to Mastery of Cognitive Tasks at Each Level (IV)

	Cognitive Task Difficulty Levels										
	≤2	3	4	5	6	7	8	9	10	11	12
Interaction Quality	-0.923*	-0.212**	0.007	-0.819***	-0.710***	-0.699*	-0.259*	-0.208***	-0.669***	-0.466**	-0.196
Home Visitor and Caregiver	(0.515)	(0.108)	(0.092)	(0.190)	(0.276)	(0.366)	(0.150)	(0.060)	(0.215)	(0.189)	(0.131)
Interaction Quality	-0.082	-0.003	-0.052**	-0.091	-0.042	0.050	-0.015	0.019*	0.004	0.015	-0.049
Home Visitor and Child	(0.130)	(0.028)	(0.021)	(0.077)	(0.068)	(0.100)	(0.055)	(0.010)	(0.042)	(0.050)	(0.045)
Teaching Ability	0.402	0.261**	-0.245**	0.770**	0.600	-0.345	0.231	0.054	0.503***	0.177	-0.205
	(0.548)	(0.101)	(0.123)	(0.342)	(0.370)	(0.434)	(0.204)	(0.060)	(0.167)	(0.274)	(0.164)
Grandmother Rearing[1]	0.032	-0.002	-0.027	0.437*	0.436**	0.006	0.283*	0.100**	0.336**	0.793***	0.117
	(0.255)	(0.062)	(0.071)	(0.258)	(0.176)	(0.225)	(0.155)	(0.045)	(0.162)	(0.209)	(0.088)
Monthly Age	-0.057***	-0.007	0.007	-0.012	-0.018	0.025	0.032*	-0.001	0.002	0.067**	-0.011
	(0.012)	(0.004)	(0.010)	(0.012)	(0.021)	(0.026)	(0.018)	(0.003)	(0.009)	(0.032)	(0.011)
Constant	2.309***	1.033***	0.970***	1.726***	2.463***	2.666*	0.516	1.025***	1.222**	-1.019	1.601***
	(0.762)	(0.259)	(0.311)	(0.540)	(0.687)	(1.541)	(0.498)	(0.095)	(0.532)	(0.884)	(0.453)
Cragg-Donald F	43.494	34.803	22.807	43.648	48.213	96.371	49.372	36.137	54.441	34.974	17.043
Kleibergen-Paap LM	65.949	62.963	43.384	53.898	55.824	89.574	72.079	54.228	90.675	34.408	52.252
Hansen J	1.962	5.604	2.901	0.858	0.779	3.639	3.913	0.754	2.669	1.392	2.451
P-Value(Hansen J)	0.375	0.061	0.234	0.651	0.678	0.162	0.141	0.686	0.263	0.499	0.294

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
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 at the previous difficulty level until the first success (inclusive) at each difficulty level by each skill type.
6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics. For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.
7. Standard errors are reported in parentheses and clustered at village level.
8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

stronger point estimates.

Appendix D reports comparable results for other skills and other measures of knowledge. The interventions have no impact on gross motor skills. In it, we also evaluate the impact of interventions on post-treatment caregiver interactions with children. We measure the frequency of the caregiver playing with the child on the tasks after each home visit.[17] The intervention only promotes the frequency of caregiver play with low-ability children.[18]

5 Nonparametric Tests of a Version of Dynamic Complementarity

Dynamic complementarity arises if early investments affect the productivity of later investments. It governs the extent to which investment at later ages can substitute for deficient early childhood investment. Heckman and Zhou (2022b) present formal tests of this proposition on our data and find evidence for it. In this section, we present some additional nonparametric evidence which does not require any particular assumption about scales of skills except the maintained assumption of comparability of knowledge *within* skill levels. We use passing rates as our measure of knowledge.

Although children enter the program at different ages, all enrolled children of the same age receive the same lesson. We determine whether late entrants can catch up.

---

[17] 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.

[18] See Table D.17.

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. Over the age range of 10-25 months, children enter the program more or less randomly with respect to age due to administrative constraints.[19]

None of the children receive training in the program before entry but may acquire skills from home instruction, imitation and maturation. Suppose that a child enters at level $\ell(s)$ for skill s at age $a^+(s,\ell)$. Some may be able to master the task from the outset, but many do not. We compute the probability of mastery for new entrants at entry age $a^+(s,\ell)$ as

$$q(s,\ell,a^+(s,\ell)) = \Pr\left(D(s,\ell,a^+) = 1\right)$$

where age appropriate lessons are administered at or near a^+. To test this, we use as new entrants children who enroll in the program who have less than one month of exposure to it. $q(s,\ell,a^+(s,\ell))$ is a measure of learning from maturation and exposure without participating in the program.

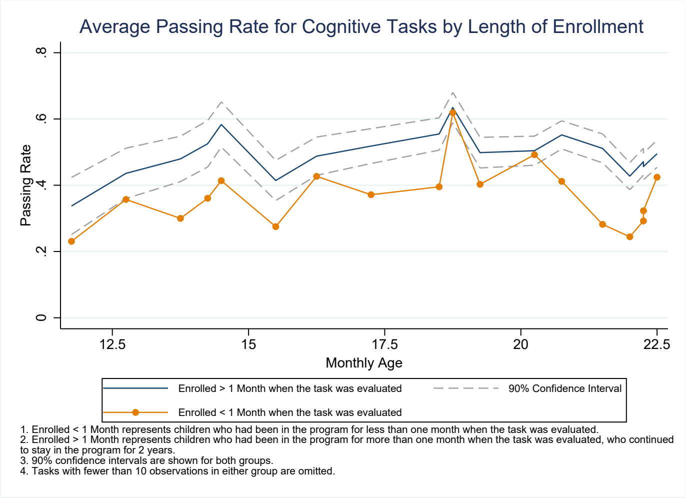
We consider performance by age at entry-level. Figure 3 shows the initial passing rate (q) for cognitive tasks by age (length of enrollment). It indicates that the knowledge of those not previously enrolled in the program (i.e., the ones who enrolled less than one month) is less than that of children at the same age who were enrolled in the program longer than one month. For most tasks, the group that is enrolled for longer than one month performs significantly better than new entrants. At the same ages, the endowments for children who just enroll in the program are smaller

---

[19] See Figure E.1.

than those for the children in for longer spells.

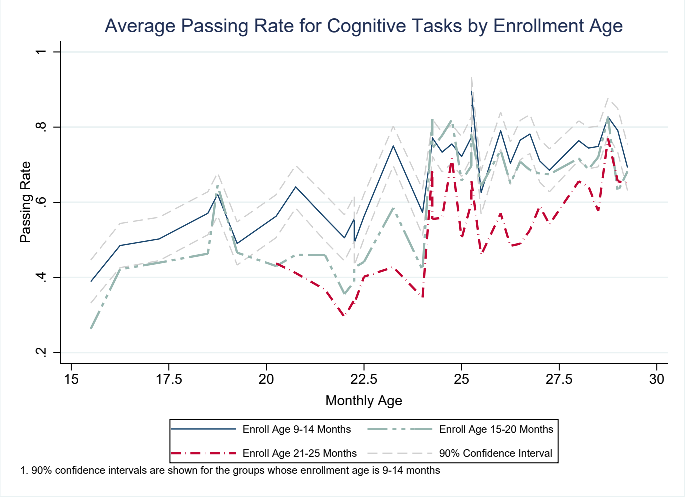
Figure 3: Cognitive Tasks Performance Comparison by Length of Enrollment



Average Passing Rate for Cognitive Tasks by Length of Enrollment

1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years.
3. 90% confidence intervals are shown for both groups.
4. Tasks with fewer than 10 observations in either group are omitted.

Figure 4: Average Passing Rate for Cognitive Tasks by Enrollment Age



Average Passing Rate for Cognitive Tasks by Enrollment Age

1. 90% confidence intervals are shown for the groups whose enrollment age is 9-14 months

Figure 4 compares the passing rates at the designated ages for different cohorts of entrants with that of those of persons who start later in the same age-specific curriculum. Consistent with dynamic complementarity, they start behind generally stay behind. The longer the child has been in the program, the higher the passing rate on cognitive tasks. There are few entrants at later ages so we trim noisy data after the age of 30 months old. Comparable patterns appear for other skills.[20]

Table 4 documents the test in detail. We ask whether those who start later catch up in terms of time to first mastery. In the vast majority of cases, they do not. In Heckman and Zhou (2022b), we show that dynamic complementarity does not operate uniformly across ability groups: normal and low-ability children display stronger dynamic complementarity effects, but high-ability children do not.

This analysis has to be qualified. If there are critical and sensitive early periods in the intervals missed by late entrants, our evidence is also consistent with that phenomena, as well as with dynamic complementarity.

To this point, we have presented empirical evidence that learning exists, the intervention boosts skill development mainly through the interactions between home visitors and caregivers, and our data are consistent with dynamic complementarity. We next develop a dynamic model to formalize these findings.

---

[20] See Appendix E.

Table 4: Cognitive Skills: Time to First Mastery by Enrollment Age

	Cognitive Difficulty Level										
	2	3	4	5	6	7	8	9	10	11	12
	Enroll (10-15) vs. (16-20)										
Mean (Age 10-15)	2.019	1.127	1.203	1.544	1.479	1.765	1.314	1.198	1.235	1.407	1.133
Mean (Age 16-20)	2.174	1.209	1.215	1.667	1.608	2.023	1.447	1.078	1.355	1.611	1.178
p-value (Age 10-15 v.s. 16-20)	0.336	**0.071**	0.830	0.368	0.216	**0.050**	0.119	**0.005**	0.309	0.190	0.628
step down *p*-value	0.834	0.447	0.850	0.834	0.755	0.363	0.588	**0.056**	0.834	0.737	0.850
N	384	268	221	442	427	431	416	249	192	167	103
	Enroll (10-15) vs. (21-25)										
Mean (Age 10-15)	2.019	1.127	1.203	1.544	1.479	1.765	1.314	1.198	1.235	1.407	1.133
Mean (Age 21-25)	1.385	1.194	1.156	2.118	2.161	2.424	1.877	1.100	1.483	1.576	1.275
p-value (Age 10-15 v.s. 21-25)	**0.002**	0.325	0.488	**0.001**	**0.000**	**0.000**	**0.000**	**0.033**	**0.047**	0.268	0.172
step down *p*-value	**0.012**	0.604	0.604	**0.008**	**0.000**	**0.001**	**0.000**	0.154	0.169	0.604	0.475
N	264	189	173	395	389	387	373	236	211	191	132
	Enroll (16-20) vs. (21-25)										
Mean (Age 16-20)	2.174	1.209	1.215	1.667	1.608	2.023	1.447	1.078	1.355	1.611	1.178
Mean (Age 21-25)	1.385	1.194	1.156	2.118	2.161	2.424	1.877	1.100	1.483	1.576	1.275
p-value (Age 16-20 v.s. 21-25)	**0.001**	0.851	0.413	**0.009**	**0.000**	**0.017**	**0.000**	0.518	0.225	0.795	0.209
step down *p*-value	**0.013**	0.956	0.873	**0.068**	**0.001**	0.106	**0.003**	0.892	0.743	0.956	0.743
N	224	141	138	385	386	384	369	293	267	240	175

1. Group (10–15) represents children whose monthly ages are between 10 and 15 at enrollment.

2. Group (16–20) represents children whose monthly ages are between 16 and 20 at enrollment.

3. Group (21–25) represents children whose monthly ages are between 21 and 25 at enrollment.

4. Time to first mastery is defined as the number of tasks a child takes until the first success (inclusive) at each difficulty level during the intervention by skill type.

5. Step down *p* values are constructed by multiple hypotheses between the earlier enrolled group and later enrolled group based on Romano and Wolf (2005a,b) .

6. Step down *p*-values are conducted by 5000 times of bootstrap.

6 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 all participants 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 a 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. Skill evolves via

$$K(a) = K(a-1) + \delta(a)\eta K(a-1) + V(Q(a)), \tag{6}$$

where η is an 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 $Q(a)$, such as family background and other investments received at home, as well as maturation effects through ages and affect the evolution of skills that operates independently of the level of $K(a-1)$. We assume skill invariance *within* each designated skill level.[21] 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 Pro-*

---

[21] The tasks within each difficulty level are essentially the same.

ductivity is $\dfrac{\partial K(a)}{\partial \delta(a)} = \eta K(a-1)$. *Static Complementarity* between skills and investment at age $a-1$ defined as: $\dfrac{\partial^2 K(a)}{\partial K(a-1)\partial \delta(a)} = \eta > 0$. *Dynamic complementarity* arises from investment at age a on the productivity of future investments is defined as $\frac{\partial^2 \boldsymbol{K(a+j+1)}}{\partial \boldsymbol{\delta(a)} \partial \boldsymbol{\delta'(a+j)}}$.[22]

Adding Shocks

A multiplicative version of the model turns out to fit the data on skill growth very well. [23] Adding i.i.d. idiosyncratic shocks in growth rates ($\varepsilon(a)$) on a log scale, skill acquisition is characterized by:

$$\ln K(a) - \ln K(a-1) \doteq \delta(a)\eta + V(Q(a)) + \varepsilon(a) \tag{7}$$

Accounting for initial conditions, Equation (7) becomes:

$$\ln K(a) \doteq \eta \sum_{j=1}^{a} \delta(j) + \sum_{j=1}^{a} V(Q(j)) + \sum_{j=1}^{a} \varepsilon(j) + \ln K(0) \tag{8}$$

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 skill levels following Rutherford (1955).

---

[22] Heckman and Zhou (2022b) show that dynamic complementarity can be affected by (a) complementarity between skills and investment in period $a+j$, (b) self-productivity (e.g., the marginal productivity of investment), and (c) the transmission of period a investment to latent skills in period $a+j+1$.

[23] Another paper, Heckman and Zhou (2022c), compares the empirical performance of multiplicative and additive models. In many aspects, the qualitative results from each are very similar but quantitative results are somewhat better for the multiplicative model as characterized by model specification tests.

Adding stochastic shocks to learning growth accounts for fadeout or acceleration off deterministic growth paths. The entire literature on fadeout of test scores (see, e.g., Duncan et al., 2022) assumes deterministic growth profiles. We allow for stochastic growth and fadeout.

Define $U(a) = \sum_{j=1}^{a} \varepsilon(j)$, a random walk, $\Delta(a) = \sum_{j=1}^{a} \delta(j)$ is cumulative lessons, and $\Lambda(a) = \sum_{j=1}^{a} V(Q(j))$. In this notation, 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 so shocks are from the same distribution independent of ability level. Conditioning on η, assumed to be independent of $U(a)$ and $K(0)$, we obtain

$$\Pr(D(a) = 1 \mid \eta, \Delta(a), \Lambda(a), K(0)) = \int_{\ln \bar{K} - \eta\Delta(a) - \Lambda(a) - \ln K(0)}^{\infty} dF(U(a)). \tag{9}$$

The General Model

Using the notation introduced in Section 3, Equation (2), the general model has the same structure 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.

Shocks at level ℓ for age a—$\varepsilon_\ell(s,a)$—are assumed to be independent across a. Their distributions may vary with ℓ and a. When estimating the model, we assume that they are i.i.d. within ℓ. $\eta(s)$ may vary by age a[24] and $\delta(a)$ captures the content

---

[24] In the estimation, η includes the interaction measures and the measure of grandmother appearance. Therefore, η changes as lessons change.

of the curriculum. Thresholds (passing standards) $\bar{K}(s,\ell)$ may also change across levels, as may $V_\ell(Q(a))$.

This model is a contribution to mathematical psychology. It unites and extends two fundamental psychometric models: the Item Response Theory (IRT) model (Lord and Novick, 1968) and the Bayesian Knowledge Tracing (BKT) model (Corbett and Anderson, 1994). The essential feature of the IRT model is captured by the threshold crossing feature (2). The BKT model is captured by the dynamics of the model (6). Unlike the BKT model, in our model knowledge $\boldsymbol{K(a)}$ is affected by education and investment is captured by $\boldsymbol{I(a)}$ so that we depart from its mechanical growth trajectory feature to account for investment that affects learning.[25]

By allowing for level-specific shocks, we account for the possibility that different difficulty levels within an assessment may have different variances and thresholds. This is indeed what we find in our estimates. We can explain "fadeout" by allowing for level-specific differences in difficulty of an assessment and level-specific responses of test takers. We next define two notions of skill invariance and show how we test for it within our model.

6.1 Testing Skill Invariance

As previously noted, there are two different interpretations of invariance of skill. The first interpretation, and the one examined in this paper, is that there exist latent skills that generate model outcomes, and they are comparable across ages and levels of skill. This assumption underlies all human capital models since Ben-Porath (1967).

---

[25] Deonovic et al. (2018) compare the IRT and BKT models and criticize them for not including investment as a determinant of learning.

The second interpretation is that there are invariant *measures* of skill comparable across skill levels (Agostinelli and Wiswall, 2022). The existence of scale-invariant measures in this sense requires invariant latent skills in the first sense.[26]

Skill invariance in the first sense assumes a common scale within and across all difficulty levels ℓ for each skill type s, although scales may vary across s. Heckman and Zhou (2022a) conduct nonparametric test of skill invariance of measures (constant units across levels) using passing rates within narrowly defined levels of tests as measures of knowledge, assuming skill invariance within the same levels. Cognitive and language aggregate Denver test scores are not the same for each group with identical knowledge as measured by passing rates within the same levels. These findings are robust across almost all the levels of knowledge. This paper develops and applies a model-based test of skill invariance in the first sense and rejects this assumption as well for most skills and most levels.

Under skill invariance in the first sense, index $K(s, \ell, a)$ cumulates *across* levels, so the measures of knowledge growth are well-defined. This requires, among other things, that in the absence of depreciation (or appreciation) associated with transitions across levels,

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

This is a property of latent variables. If all components of the technology of skill

---

[26] Agostinelli and Wiswall define invariant measures in the following way. Let $K_i(a)$ be child i's human capital in the first sense. Let $Z_i(m, a)$ be child i's score on a measure of $K_i(a)$ at age a. Measured age invariance is $E(Z_i(m, a) \mid K_i(a) = \tau) = E(Z_{i'}(m, a+t) \mid K_{i'}(a+t) = \tau)$; $i \neq i'$, all t. This is a property of a *measure* of a latent skill, assumed to be age invariant in the first sense.

formation (Equation (1)) shift across levels, the hypothesis of skill-invariant scales lacks testability because the scale is not directly observed, and technology parameters can be redefined to impose invariance. Some parameters must be invariant across levels to conduct this test, although they need not necessarily be the same parameters across all levels. We test for skill invariance in the first sense, maintaining the assumption of invariance *within the same levels*. Our proof of model identification in Appendix F makes this point precise. Note that the assumed lack of depreciation (or appreciation) is only a property at boundaries. There can be either (or both) at work in interior segments.

If scales change across levels, but human capital scales are somehow connected, it follows that

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

where Γ_ℓ is a general function. If there is total depreciation of skills in transitions from $\ell - 1$, Γ_ℓ is the zero function. Skill invariance in the first sense at $\ell - 1$ sets $\Gamma_\ell = I$, the identity function. In this paper, we consider only affine transformations for $\Gamma_\ell(\cdot)$:

$$\Gamma_\ell(K(s, \ell, \underline{a}(s, \ell))) = \gamma_{0,\ell} + \gamma_{1,\ell}(K(s, \ell - 1, \bar{a}(s, \ell - 1))). \tag{10}$$

Setting $\gamma_{0,\ell} = 0$ and $\gamma_{1,\ell} = 1$ captures the notion of skill invariance. More general transformations are admissible but we use the affine transformation as a first order linear approximation of the general function.

We now present the intuition for how we can test for skill invariance in the first sense. We do not have direct measures of latent skills. Instead, we 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).

6.2 Model 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 F. We use means and covariances because we assume normal errors in estimation. However, drawing on Heckman and Vytlacil (2007) and Matzkin (1992, 2007), we show in Appendix F that we can nonparametrically identify the joint distributions of unobserved variables up to normalizations 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 (8) to read:

$$\ln K(1,a) = \underbrace{\eta \sum_{j=1}^{a} \delta_1(j)}_{\text{learning}} + \underbrace{V_1(a)}_{\substack{\text{maturation and}\\ \text{exposure effects}}} + \underbrace{U_1(a)}_{\text{shocks}} + \ln K(0), \tag{11}$$

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 $\delta_1(a)\eta(X) = \bar{\beta}_1(X) + \omega$, where X are covariates including ability and interactions. $X \perp\!\!\!\perp [\omega, \varepsilon_1(j)]$ for all j. ω is an individual-specific random shock, with $E(\omega) = 0$, and $\omega \perp\!\!\!\perp (\Upsilon, \varepsilon_1(j))$ for all j. It captures heterogeneity in learning ability. 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 ω_j.[27] We assume that the learning component $\delta_1(a)$ is constant within each level but can differ across levels. $V_1(a)$ is shorthand for $\sum_{j=1}^{a} V_1(Q(j))$.

Equation (11) can be rewritten in the notation for the general case allowing for heterogeneity in $\ln K(0)$:

$$\ln K(1,a) = \mu_1 + \mu_0(Z) + V_1(a) + \bar{\beta}_1(X)a + \underbrace{\left\{ a\omega + \sum_{j=1}^{a} \varepsilon_1(j) + \Upsilon \right\}}_{\Psi_1(a)} \tag{12}$$

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

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

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

where $\mu_1^* = \mu_1 - \bar{K}(1)$ and μ_1 collects any other model intercepts. If any slope 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 μ_1^*, $\mu_0(Z)$, $\bar{\beta}_1(X)$, $V_1(a)$ up to scale. Since we can identify the ratio of $\frac{\sigma(1,j)}{\sigma(1,j')}$, $\sigma(1,a)$, $\sigma(1,a')$ are

---

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

identified up to a normalization (e.g., $a, a' \neq j$) (see Heckman, 1981 and Heckman and Vytlacil, 2007).

Using the definition of $\sigma^2(1,a) := a^2\sigma^2_\omega + a\sigma^2_{\varepsilon(1)} + \sigma^2_\Upsilon$, we have the following equations:

$$\begin{aligned}
\sigma^2(1,a) &= a^2\sigma^2_\omega + a\sigma^2_{\varepsilon(1)} + \sigma^2_\Upsilon \\
\sigma^2(1,a') &= (a')^2\sigma^2_\omega + a'\sigma^2_{\varepsilon(1)} + \sigma^2_\Upsilon \\
\sigma^2(1,j) &= j^2\sigma^2_\omega + j\sigma^2_{\varepsilon(1)} + \sigma^2_\Upsilon .
\end{aligned}$$

In these equations, the left-hand sides are identified up to the scale (i.e., $\sigma^2(1,j) = 1$). On the right-hand sides, there are three unknown terms σ^2_ω, $\sigma^2_{\varepsilon(1)}$, and σ^2_Υ. When $a \geq 3$ (i.e., three different tasks at level one), we can identify all three terms: σ^2_ω, $\sigma^2_{\varepsilon(1)}$, and σ^2_Υ with sufficient variation in a and j.

Adopting a similar notation for levels $\ell > 1$, if we assume skill 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 level 2 at age a can be written as:

$$\begin{aligned}
\ln K(2,a) = \mu_2 + V_2(a) + \bar{\beta}_2(X)(a - \bar{a}(1)) + \sum_{j=\underline{a}(2)}^{a} \varepsilon_2(j) + \ln K(1, \bar{a}(1)) \\
= \mu_1 + \mu_2 + \mu_0(Z) + V_1(\bar{a}(1)) + V_2(a) + \bar{\beta}_2(X)(a - \bar{a}(1)) + \bar{\beta}_1(X)\bar{a}(1) \\
+ \underbrace{\left\{ \sum_{j=\underline{a}(2)}^{a} \varepsilon_2(j) + (a - \bar{a}(1))\omega + \sum_{j=1}^{\bar{a}(1)} \varepsilon_1(j) + \bar{a}(1)\omega + \Upsilon \right\}}_{\Psi_2(a)} . \qquad (13)
\end{aligned}$$

Given the initial normalization at level one (i.e., $\sigma(1,j)=1$) and identification of the parameters in the first level (up to scale), we can identify $V_2(a)$ and $\bar{\beta}_2(X)$ up to scale $\sigma(2,a)$, where

$$\begin{aligned}
\Psi_2(a) = \sum_{j=\underline{a}(2)}^{a} \varepsilon_2(j) + (a-\bar{a}(1))\omega + \sum_{j=1}^{\bar{a}(1)} \varepsilon_1(j) + \bar{a}(1)\omega + \Upsilon \\
\sigma^2(2,a) :=& \mathrm{Var}\Psi_2(a) \\
\mathrm{Var}\Psi_2(a) =& \sigma^2_\Upsilon + a^2\sigma^2_\omega + (a-\underline{a}(2))\sigma^2_{\varepsilon(2)} + \bar{a}(1)\sigma^2_{\varepsilon(1)}.
\end{aligned}$$

Notice that we have identified σ^2_ω, $\sigma^2_{\varepsilon(1)}$, and σ^2_Υ, and the only term not identified in $\mathrm{Var}\Psi_2(a)$ is $\sigma^2_{\varepsilon(2)}$. We now discuss how to identify this term. Consider the covariance term $\mathrm{Cov}\left(\frac{\Psi_2(a)}{\sigma(2,a)}, \frac{\Psi_2(a')}{\sigma(2,a')}\right)$

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

In the equation just written, we observe the left-hand side value. On the right-hand side, the only unknown term is the variance of shocks at level 2 (i.e., $\sigma^2_{\varepsilon(2)}$). Therefore, we can identify the value of $\sigma^2_{\varepsilon(2)}$. After identifying $\sigma^2_{\varepsilon(2)}$, we can identify the scale of variance term $\sigma^2(2,a)$. Then, we can identify $V_2(a)$ and $\bar{\beta}_2(X)$ up to $\sigma(2,a)$.

From the previous discussion, we can identify, for all $\ell \geq 2$, the variance $\sigma(\ell,a)$ without imposing additional normalization at levels ℓ ($\ell \geq 2$). The only normalization we need is on the scale of variance term $\sigma(1,j)=1$ at level one.

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 and a' in the appropriate intervals and the technologies at each level subject to the initial normalization. Details concerning nonparametric identification are discussed in Appendix F. We do not develop this point further because we adopt 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 at level one.

6.2.1 Testing the Skill Invariance Assumption

Under skill invariance characterized by Equation (10) with $\gamma_{0,\ell} = 0$ and $\gamma_{1,\ell} = 1$, we obtain tight restrictions on the coefficients across levels. Relaxing scale invariance adds two new parameters $(\gamma_{0,2}, \gamma_{1,2})$ to Equation (13):

$$\ln K(2,a) = \gamma_{0,2} + \mu_2 + V_2(a) + \bar{\beta}_2(X)(a - \bar{a}(1)) + \sum_{j=\underline{a}(2)}^{a} \varepsilon_2(j) + \gamma_{1,2} \ln K(1, \bar{a}(1)).$$

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

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

From the previous analysis, the term in braces is identified up to the previously stated 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 ℓ. As before, we can estimate $\ln \bar{K}(2)$ up to scale net of intercepts, and we can identify the scale.

7 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 G. The moments used in forming the estimates are presented in Table H.1. The model passes goodness of fit tests (see Appendix H).[28] Appendix H also plots model predictions vs data for each skill, with and without skill invariance[29]. In general, imposing the age invariance assumption produces worse fits, a point developed further below. We report estimates in the text that do not impose skill invariance. Estimates imposing skill invariance are presented in Appendix I.

7.1 Estimates

We report our empirical results by skill level.

---

[28] When we separate estimates by gender, we find no differences in the structural parameters. The initial conditions favor girls and that explains their better scores on the tests. See Zhou et al. (2022).

[29] See Figures H.1, H.7, and H.13 for language, cognition, and fine motor skills, respectively.

7.1.1 Language Skills

Figure 5a displays estimates of the estimated minimum skill level requirement for each level. This is defined relative to $\bar{K}(1)$, assuming no shift in model intercepts for each skill across levels apart from that due to skill accumulation. As expected, the skill level required to pass tasks is monotonically increasing across difficulty levels. We do not impose any restriction on the order of the $\bar{K}(\ell)$. The estimates show that on average the difficulty levels in the curriculum are consistent with child task performance. The variances of shocks at each level display different patterns, reflecting differentials in ability. Figure 5b presents estimates of the variances. 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 6, 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 6). Level-specific shocks can intrude to alter the monotonicity delivered by the deterministic model and to capture the lack of fit of the model to the data.[30]

Note that "fadeout" as measured by passing rates appears within level 11 and across levels 6-11 as a consequence of patterns of item difficulties and variances. This despite the stochastically monotonic increase in knowledge.

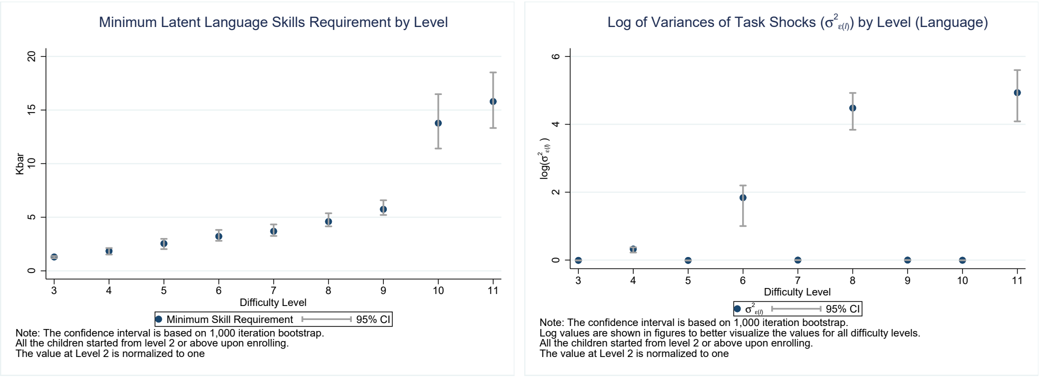
7.1.2 Cognitive Skills

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

---

[30] See Figure H.1b.

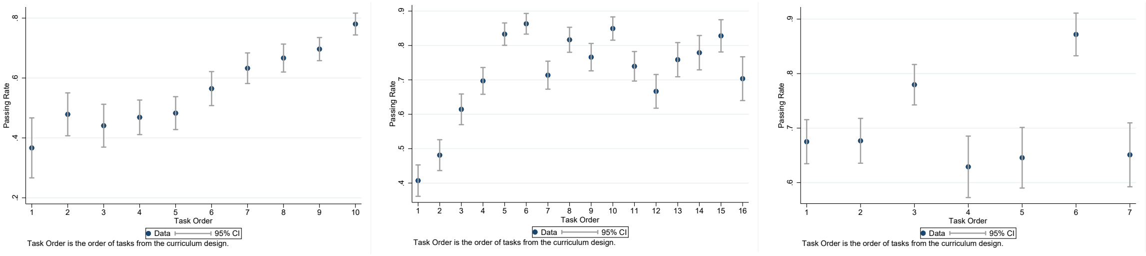
Figure 5: Language Skill



(a) $\bar{K}(\ell)$

(b) $\sigma^2_{\varepsilon(\ell)}$

Figure 6: Average Passing Rate of Language Tasks: $p(s, \ell)$ (Raw Data)



(a) Level 6

(b) Level 8

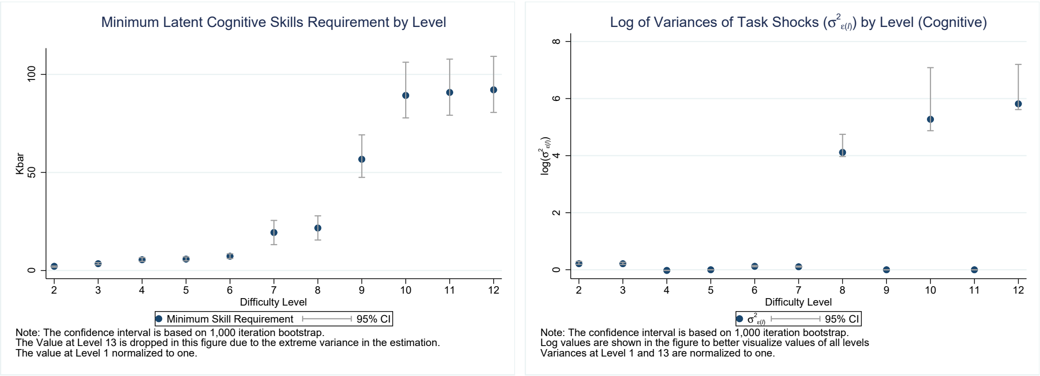
(c) Level 11

levels, thus explaining “fadeout” even when, on average, skill levels are increasing.

7.1.3 Fine Motor Skills

A similar pattern arises for fine motor skills.

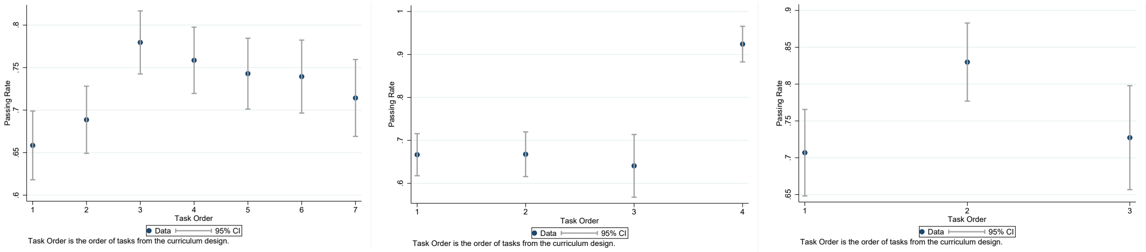
Figure 7: Cognitive Skill


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

Note: The confidence interval is based on 1,000 iteration bootstrap.
The Value at Level 13 is dropped in this figure due to the extreme variance in the estimation.
The value at Level 1 normalized to one.

Note: The confidence interval is based on 1,000 iteration bootstrap.
Log values are shown in the figure to better visualize values of all levels
Variances at Level 1 and 13 are normalized to one.

Figure 8: Average Passing Rate of Cognitive Tasks: $p(s,\ell)$ (Raw Data)

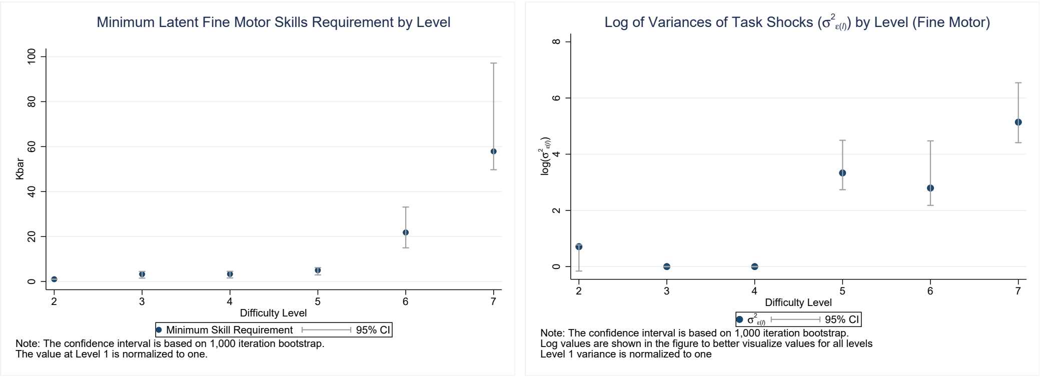

(a) Level 8 (b) Level 10 (c) Level 12

Task Order is the order of tasks from the curriculum design.

Task Order is the order of tasks from the curriculum design.

Task Order is the order of tasks from the curriculum design.

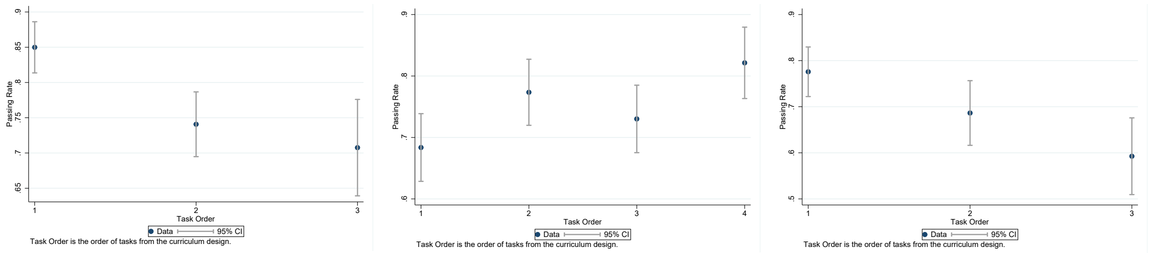
Figure 9: Fine Motor Skill


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

Note: The confidence interval is based on 1,000 iteration bootstrap.
The value at Level 1 is normalized to one.

Note: The confidence interval is based on 1,000 iteration bootstrap.
Log values are shown in the figure to better visualize values for all levels
Level 1 variance is normalized to one

Figure 10: Average Passing Rate of Fine Motor Tasks: $p(s,\ell)$ (Raw Data)



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

Task Order is the order of tasks from the curriculum design.

7.2 Learning Components and Task Performance

In this section, we examine how the learning component in our structural model $\delta_\ell E(\eta)$ explains child task performance. The δ_ℓ term captures the curriculum content at each difficulty level, which is common across all children, and the $\eta(X)$ term includes interaction quality measures between home visitors and caregivers/children, home visitors' teaching quality, and grandmother rearing during the intervention.

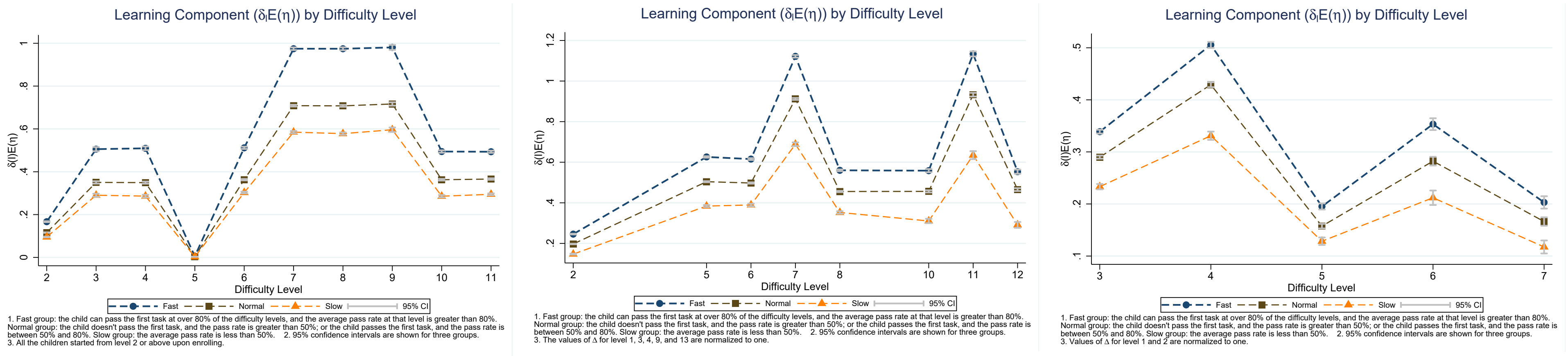
The intervention interaction variables (entered as X in $\beta_\ell(X)$) are significant determinants of child learning for each task. This finding is consistent with the results reported in Section 4. The interaction between the home visitor and the caregiver is the only consistently positive interaction that promotes skills (see Appendix I).[31] The grandmother, as the main caregiver, often has significantly negative effects on learning.[32]

---

[31] All the estimation results are presented in Appendix I.

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

Figure 11: Estimates of $\delta_{(}\ell)E(\eta)$ by Ability Group



(a) Language

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. 3. All the children started from level 2 or above upon enrolling.

(b) Cognitive

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. 3. The values of Δ for level 1, 3, 4, 9, and 13 are normalized to one.

(c) Fine Motor

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. 3. Values of Δ for level 1 and 2 are normalized to one.

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

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 11). We also find that higher caregiver education levels are significantly associated with better language skills when children are first enrolled in the program (see Table I.1). There is substantial learning for children exposed to more educated mothers.

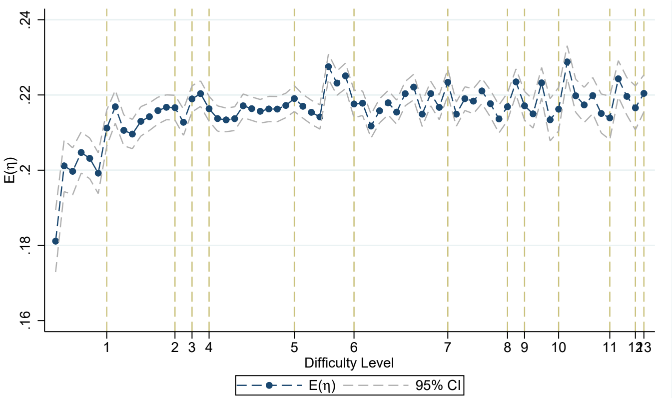
We now focus on how the $\eta(X)$ term affects child performance on tasks. Figure 12a shows the mean of $\eta(X)$ for each cognitive task. We identify it using β_ℓ and normalize $\delta(1) = 1$. There is an increasing pattern of $E(\eta)$ within difficulty levels. In Figure 12b, we break down the estimated $E(\eta)$ values by ability group.[33] Children in the normal ability group contribute the most growth in learning. Children in the fast group master the task quickly, usually on the first try. Thus, they have little subsequent learning growth 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 pattern 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 other skills.[34] For fine motor tasks, there is a similar pattern for tasks greater than 4, although learning is not substantial at any level.

---

[33] Fast group: the child passes the first task at over 80% of the difficulty levels, and the average passing rate at that level is greater than 80%. Normal group: the child does not pass the first task, and the passing rate is greater than 50%; or the child passes the first task, and the passing rate is between 50% and 80%. Slow group: the average passing is less than 50%.

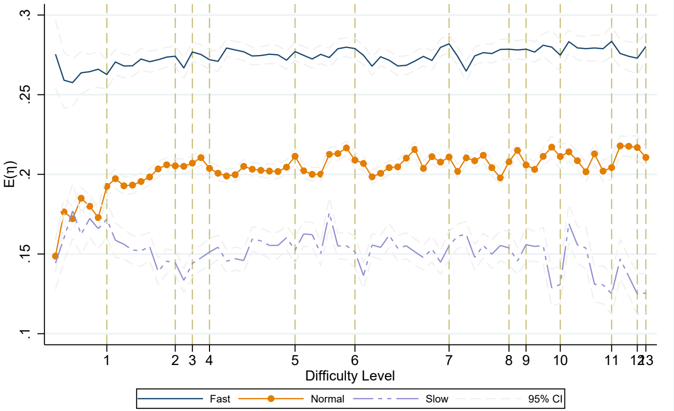
[34] See Figures J.1-J.4 in Appendix J.

Figure 12a: 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 12b: 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.

Appendix Tables J.1-J.3 compare each interaction component by family education background, child ability category, and age of enrollment. As expected, the interaction quality between the home visitor and caregiver contributes the most to the learning component η. The interaction quality between the home visitor and the caregiver is higher for the household with higher family education levels. Also, the interaction quality measures are significantly different by ability groups and age of enrollment.

7.3 Testing Skill Invariance in the First Sense

Under our parameterization of skill invariance in the first sense, $\gamma_{1,\ell} = 1$. Note that $\gamma_{1,\ell} = 1$ implies the validity of a constant-unit latent skill across ℓ and $\ell - 1$. Figure 13a shows that estimates of $\gamma_{1,\ell}$ for each skill level for models estimated without imposing the restriction $\gamma_{1,\ell} = 1$. Table 5 shows the χ^2 test results for each level and skill. Our estimates partially support skill invariance. For language and cognitive skills, at some levels, skill invariance cannot be rejected. For example, we cannot reject skill invariance for language skills between levels 8-11 (i.e., 8-9, 9-10, and 10-11).[35] However, it is decisively rejected at levels 4-6. Table 6 lists the task content for difficulty levels 8-11; it shows that the task content is very similar across these different levels. However, the null hypothesis of skill invariance across all levels is rejected. 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.

---

[35] $\gamma_{1,\ell} = 1$ implies a uniform scale for latent skill variables between level ℓ and level $\ell - 1$. For example, the coefficient at level 8 for language skills (i.e., 0.562) presents the scale between level 7 and level 8.

Figure 13: Tests of the Null Hypothesis of Skill Invariance in the First Sense

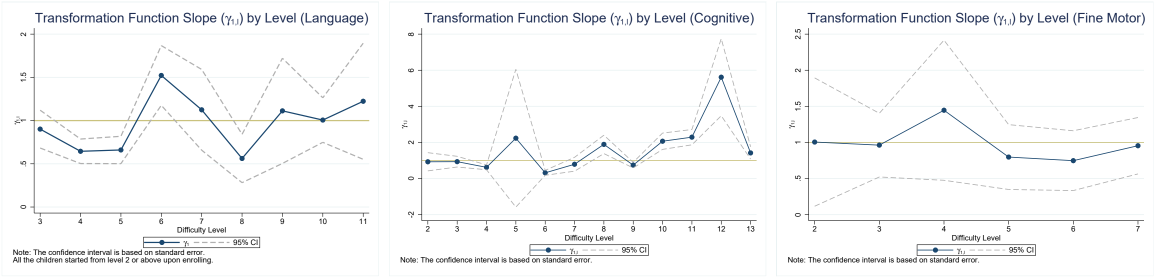

(a) Language Skill (b) Cognitive Skill (c) Fine Motor Skill

Table 5: Skill Invariance Hypothesis Tests by Levels

	Language			Cognitive			Fine Motor		
	Slope($\gamma_{1,\ell}$)	$\chi^2(\cdot)$	p-value	Slope($\gamma_{1,\ell}$)	$\chi^2(\cdot)$	p-value	Slope($\gamma_{1,\ell}$)	$\chi^2(\cdot)$	p-value
Level 2				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		44.051	0.000		71.398	0.000		2.827	0.830

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.

Table 6: 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.

We also test for skill invariance in the first sense for cognitive and fine motor skill tasks. Similarly, we reject the null of age invariance across all the levels of the cognitive skill tasks. However, we find evidence in support of skill invariance for fine motor skill tasks, which mainly test drawing skills.

In sum, our estimates do not support skill invariance in the first sense across all levels for both language and cognitive skills, but the assumption cannot be rejected for some levels and some skills. For example, we cannot reject skill invariance between levels 8, 9, and 10 for tasks testing language skills. Skill invariance appears to be a valid description of fine motor skills at all levels. Our findings call into question standard practice that relies on skill invariant measures for analyzing skill growth and value-added.

Our evidence on skill invariance in the first sense 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. That is a task left for the future.

8 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, as documented in Zhou, Heckman, Wang, and Liu (2022). We investigate the impacts of different types of interactions on child achievement measures: interactions between home visitors and caregivers, interactions between home visitors and children, the quality of the teacher, and the frequency of the caregiver playing with the child after the class. High-quality interactions between the home visitor and the caregiver significantly improve child skill development in multiple dimensions, but the other features of the program are not generally effective. We find evidence consistent with dynamic complementarity using methods that do not rely on arbitrary measures of skills.

We develop and estimate a dynamic learning model to rationalize our evidence on program impacts. The model captures patterns of learning in our data and explains how skills evolve at weekly levels. We measure the growth in knowledge across difficulty levels. Our model explains the frequently noted phenomenon of "fadeout" as a consequence of the stochastic nature of learning and the variation in performance across skill assessments. We introduce learning through investment and stochastic shocks into the standard IRT and BKT models of psychometrics.

High-ability children start strong and their knowledge generally does not improve within levels because they generally master tasks on the first attempt. Normal-ability children learn more but they have more to learn. Low-ability children also learn, but very slowly. Parental play accelerates their learning, but not that of children of other ability levels. Going forward, in designing the program, adaptive lessons that accelerate high-ability children will promote greater learning for high-ability children.

We formally test whether skill invariance in the sense of existence of a constant units scale holds across skill levels. We find evidence supporting such skill invariance for certain skills at certain difficulty levels, but we reject the assumption as a global characterization, except for fine motor skills. This finding calls into question standard practice that assumes the existence of invariant measures for analyzing child development and the value added of teachers and schools. This evidence is in accord with the findings of Heckman and Zhou (2022a) showing the nonexistence of invariant measures of skills across levels.

There is clearly room for improvement in our research. Allowing for the cross-productivity of different skills in shaping the growth of skills, following Cunha et al. (2010) is an obvious and important extension left for the future. So is semiparametric implementation of the model.

References

Agostinelli, F. and M. Wiswall (2022). Estimating the technology of children's skill formation. NBER Working Paper No. 22442, Accepted at *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 (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.

Bai, Y. (2022). Optimality of matched-pair designs in randomized controlled trials. *Conditionally accepted by the American Economic Review.*

Bai, Y., J. P. Romano, and A. M. Shaikh (2021). Inference in experiments with matched pairs. *Journal of the American Statistical Association.*

Ben-Porath, Y. (1967, August). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*(4, Part 1), 352–365.

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*.

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

Cawley, J., J. J. Heckman, and E. J. Vytlacil (1998). Cognitive ability and the rising return to education. NBER Working Paper 6388, National Bureau of Economic Research.

Corbett, A. T. and J. R. Anderson (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction 4*(4), 253–278.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

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.

Deonovic, B., M. Yudelson, M. Bolsinova, M. Attali, and G. Maris (2018). Learning meets assessment. *Behaviormetrika 45*(2), 457–474.

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

Duncan, G., A. Kalil, M. Mogstad, and M. Rege (2022). Investing in early childhood development in preschool and at home. In L. Woessmann, E. Hanushek, and S. J. Machin (Eds.), *Handbook of the Economics of Education*, Volume 6. North-Holland. Forthcoming.

Freyberger, J. (2021). Normalizations and misspecification in skill formation models.

García, J. L. and J. J. Heckman (2022). Parenting promotes social mobility within and across generations. Under Review, *Annual Reviews*.

Gertler, P., J. J. Heckman, R. Pinto, S. M. Chang, S. Grantham-McGregor, C. Vermeersch, S. Walker, and A. S. Wright (2022). Effect of the Jamaica early childhood stimulation intervention on labor market outcomes at age 31. NBER Working Paper 29292. Under Revision.

Gertler, P., J. J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. 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. and J. Zhou (2022a, April). Measuring knowledge and learning. Working Paper 29990, National Bureau of Economic Research.

Heckman, J. and J. Zhou (2022b). Nonparametric tests of dynamic complementarity. Unpublished manuscript, University of Chicago.

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 de l'inséé 30/31*, 227–269.

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.

Heckman, J. J. and J. Zhou (2022c). Comparing multiplicative and additive specifications for models of reinforcement learning. Under Preparation, University of Chicago.

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.

List, J. A. (2020). Non est disputandum de generalizability? A glimpse into the external validity trial. NBER Working Paper 27535.

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

Lu, B., R. Greevy, X. Xu, and C. Beck (2011). Optimal nonbipartite matching and its statistical applications. *The American Statistician 65*(1), 21–30.

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.

Romano, J. P. and M. Wolf (2005a, March). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*(469), 94–108.

Romano, J. P. and M. Wolf (2005b). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

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. M. Hunt (1975). *Assessment in Infancy: Ordinal Scales of Psychological Development*. Urbana, Illinois: University of Illinois Press.

van der Linden, W. J. (2016). *Handbook of Item Response Theory: Volume 1: Models*. Boca Raton, FL: 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 (2022). The impacts of a prototypical home visiting program on child skills. Working Paper 27356, National Bureau of Economic Research.

Zhou, J., J. Heckman, F. Wang, and B. Liu (2022). Early childhood learning patterns for a home visiting program in rural China. *Journal of Community Psychology*.

Zhou, J., J. J. Heckman, B. Liu, M. Lu, S. M. Chang, and S. Grantham-McGregor (2022). Comparing China REACH and the Jamaica home visiting program. Revised and Resubmitted, *Pediatrics*.

Interactions as Investments: The Microdynamics and Measurement of Early Childhood Learning APPENDIX[*]

James Heckman and Jin Zhou

Center for the Economics of Human Development
and Department of Economics
University of Chicago

December 24, 2022

---

[*] 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), the Bonn IZA-Briq Life-Cycle Workshop (August 19, 2021), and at Rice University (November, 2021). We have benefited from the comments of the editor, John List, and four insightful referees, as well as those of our discussant Lance Lochner at a September 2019 Becker Friedman Institute conference in Chicago, as well as those of Flavio Cunha, and Pat Kyllonen. Fuyao Wang, Haihan Tian, and Alejandra Campos contributed highly competent and insightful research assistance and commentary.

Contents

A 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 progression 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 the children's multidimensional skills. Based on the main content of tasks, the tasks in the curriculum cover four domains of skills.[1] The categories help researchers understand how the main type of skills develop based on the measures in the curriculum. Next, we document all the tasks in the China REACH curriculum by four domains of skill types. 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.

A.1 Skills Taught in the Curriculum

Fine motor, gross motor, language, and cognitive skills are taught. Within each skill group, based on the content in the skills, skills are ordered by difficulty level following the patterns developed by Palmer (1971). Skills are sorted into different

---

[1] We also 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.

difficulty levels. For example, for fine motor drawing lessons, there are seven difficulty levels.[2,3] In general, the higher difficulty level for 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 the tasks into different difficulty levels are categorized.

For example, Fine Motor Drawing lessons focus on child's ability to use a writing utensils with increasing skills. First, a child is asked to hold the utensil to make markings. Next, the child should 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 A.1.)

Table A.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 circle, vertical line, and horizontal line
Level 5	Draw circle, many lines, and cross 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 difficulty level category

---

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

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

descriptions are listed in this section. The number of lessons within each difficulty level depends on the content in the curriculum. For example, there are six assessments at difficulty level 3 for fine motor drawing skills and only 2 assessments at difficulty level 2.

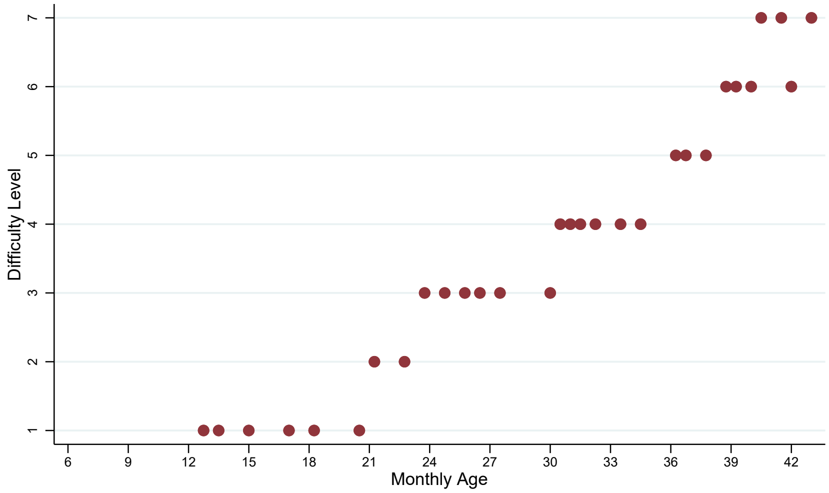
Figure A.4 gives the timing of each fine motor drawing assessments appearing in the curriculum design. For the designated skills, difficulty level 1 covers from 12 months and 3 weeks to 20 months and 2 weeks. This timing means that when the child is 12 months and 3 weeks old, the home visitor will teach her the first fine motor drawing skill. When she is 20 months and 2 weeks old, the home visitor will teach her the 6th 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 A.4, when fine motor lessons at difficulty 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 A.2, such as the child's proficiency in understanding the objects. Seventy-four lessons are sorted into the listed 13 ordered difficulty levels.[4] It covers the process of how the child learns to know an object and

---

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

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



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

understand the function of the object.

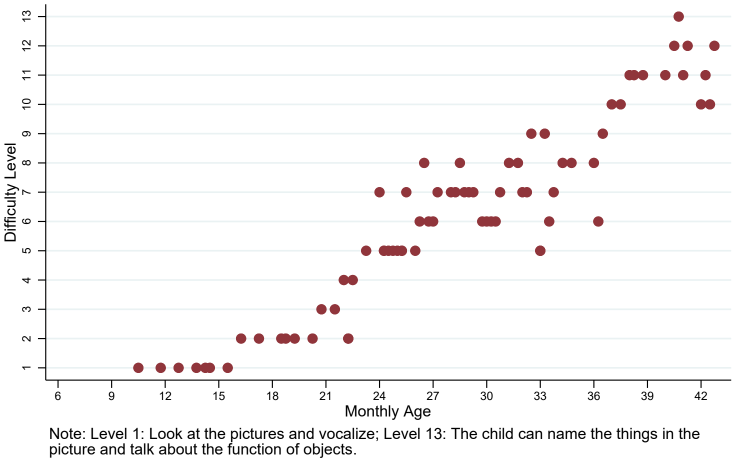
The lessons in the cognitive knowledge of objects unit progress from a simple understanding of the concept of pictures by acknowledging with vocalizations, to using receptive (heard) language to identify certain pictures. Receptive language is a skill developed prior to the expressive language where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

Table A.2: Difficulty Level List for the Cognitive Understanding Objects 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 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 A.2 shows the timing of each cognitive (knowing objects and understanding the object's function) level in the curriculum. The number of lessons varies across difficulty levels according to the curriculum content itself. Table A.3 presents detailed information about the six lessons (and assessments) that are labeled as difficulty level one directed to ten-month to 15 month-old curriculum content. In Table A.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.

Figure A.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.

Table A.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 at different skill levels for multiple levels of skill at each weekly age. For each type of skill, the task difficulty levels based on the content of the tasks and the guideline of Uzgiris and Hunt (1975) and Palmer

(1971) are constructed. 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 comparable across children. We use achievement at each level of skill as our measure of knowledge.

A.2 Fine Motor Task

Fine motor skill involves finger movements, such as the ability to grasp, release and stitch; and drawing and writing skills. Here we consider two types of fine motor skills: (1) finger movements related to grasping, releasing, stitching; (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 the ability to write letters or words.

The first category is related to finger movements regarding grasping, releasing, stitching.[5] In Table A.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.

---

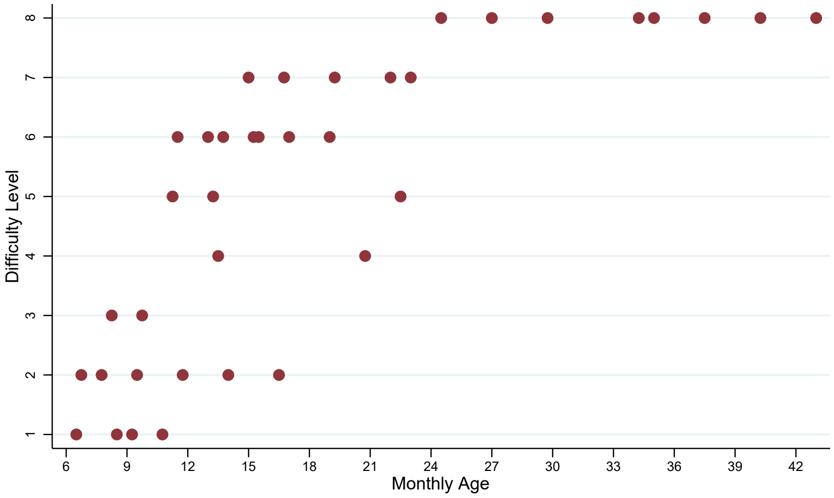
[5] These milestones are justified at `https://www.chrichmond.org/therapy-services/occupational-therapy/developmental-milestones/fine-motor-skills-birth-to-2-years` and `http://www.kamloopschildrenstherapy.org/fine-motor-skills-infant-milestons`.

Table A.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

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

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



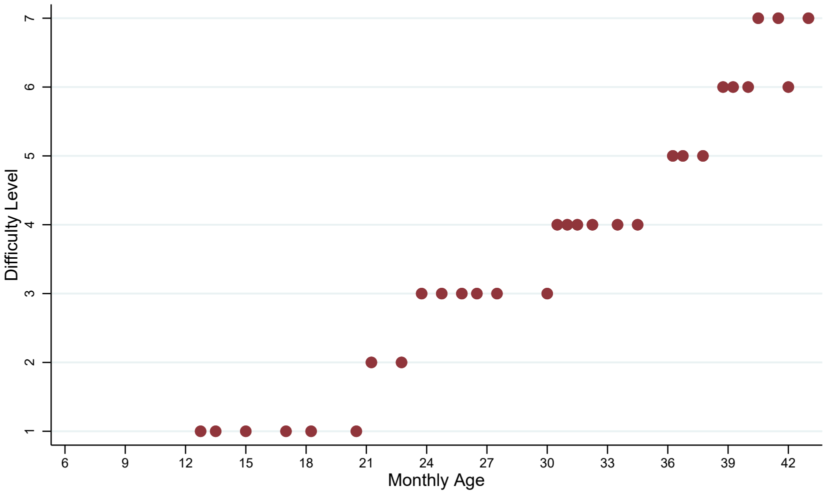
Note: Level 1: rattle; Level 8: String beads.

The second category is related to drawing and manual writing ability. The fine motor drawing tasks in Table A.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 A.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 circle, vertical line, and horizontal line
Level 5	Draw circle, many lines, and cross lines
Level 6	Draw a cross (or T), curves and zigzag curves
Level 7	Draw caterpillars

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



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

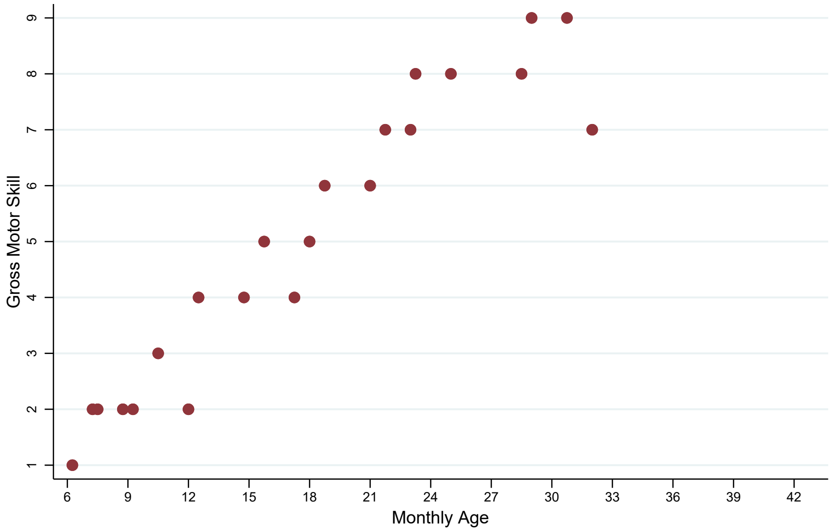
A.3 Gross Motor Skill

Gross motor skill is any skill that requires movement and precision of large muscles in the body. 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 A.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 A.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.

A.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.

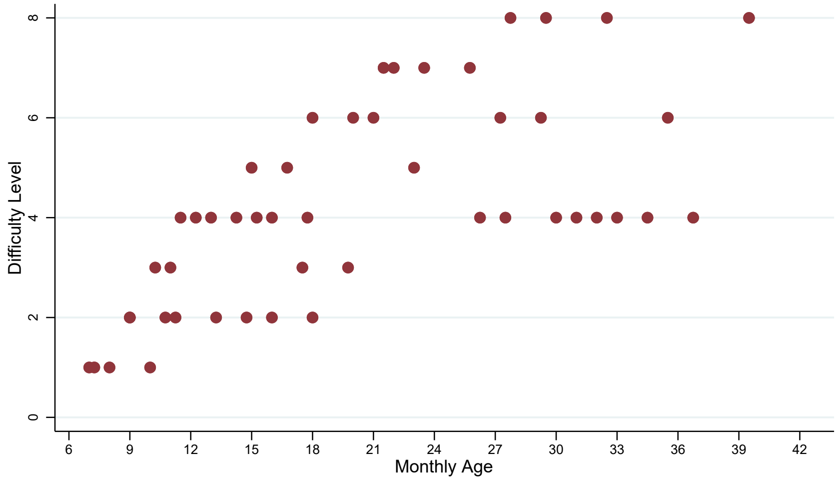
A. Spatial Skill:

Spatial skills rely 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 A.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 A.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. Knowing objects and objects’ functions:

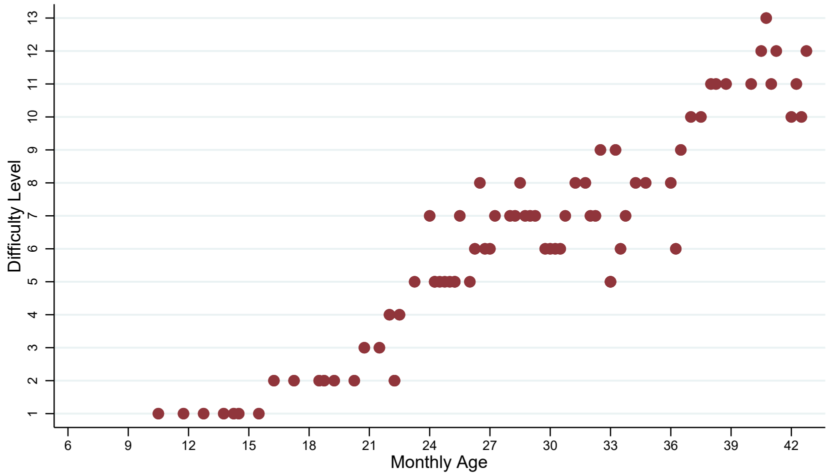
The knowing objects task set introduces preliteracy skills. It involves progressing interaction with pictures of objects and elements of storytelling. The tasks in the

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

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 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 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 A.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.

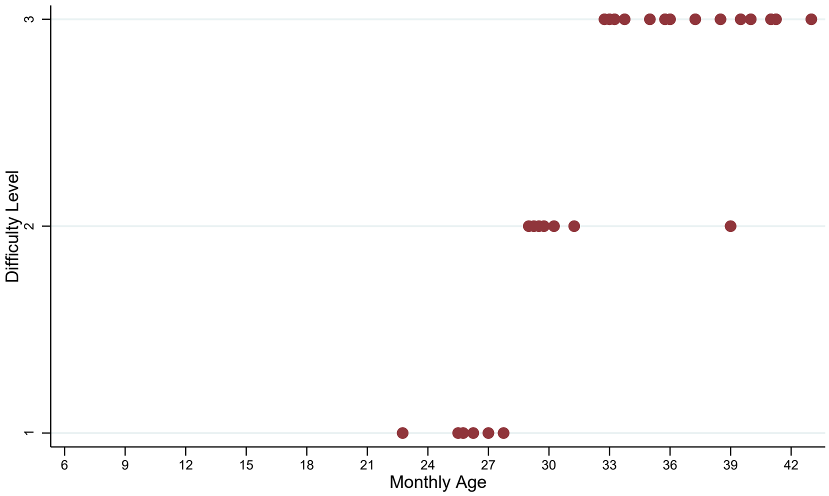
C. Color:

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 A.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	Match different colors

Figure A.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 A.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”, “behind”

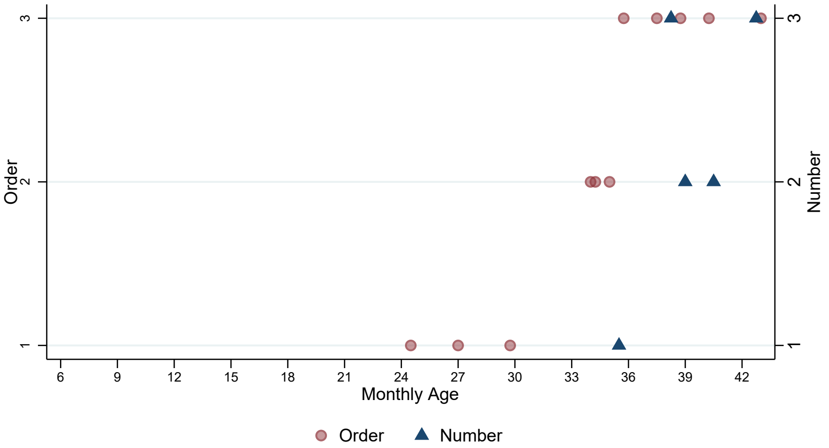
Cognitive ability progresses into more abstract concepts of direction “upward” and “downward.” Then, relative concepts of “first” “last” or “behind” are introduced.

Table A.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 wrote 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 A.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.

F. Match: match different pieces from simple puzzles to complicated puzzles

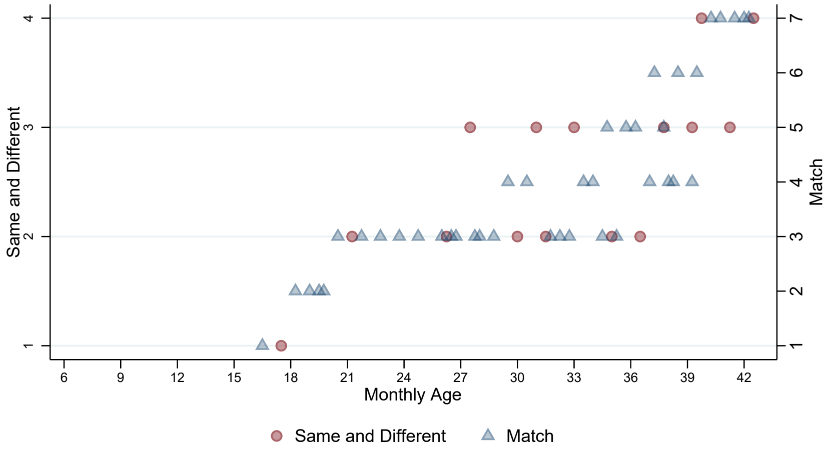
This set of tasks builds on the child's spacial awareness skills. The ability to fill in missing objects and to understand how objects fit together is important in the development of spacial awareness. The individual tasks progress from simply placing

1-2 puzzle pieces, to completing the puzzle, to making patterns and using emerging language skills to describe pieces. As the children gain proficiency in these skills, they are able to complete puzzles of increasing complexity, as well as restoring the jumbled pieces to the original puzzle.

Table A.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 A.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.

A.5 Language Skill

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

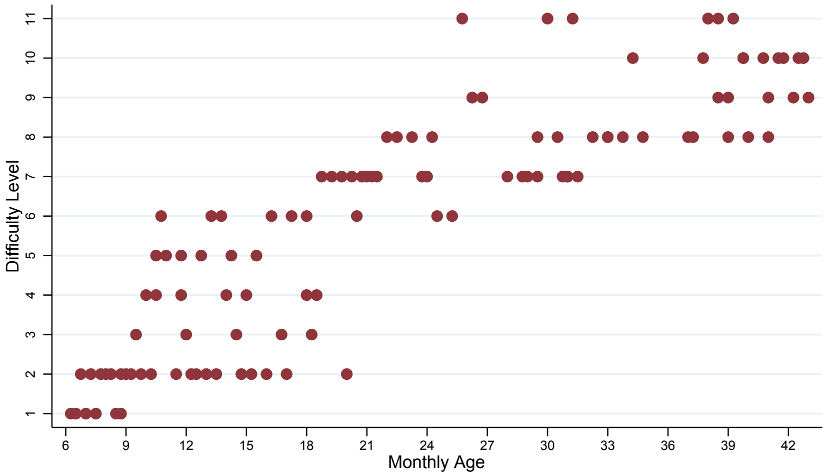
A. Learn words

Table A.13: Difficulty Level List for Language (Knowing objects and Understanding the function of objects) 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. The language skills tasks begin simply 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 picture and learn to describe them. Finally, the child will learn the names and uses of objects in the child's everyday environment.

Figure A.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. Dialogue: caregiver talks to children

Table A.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

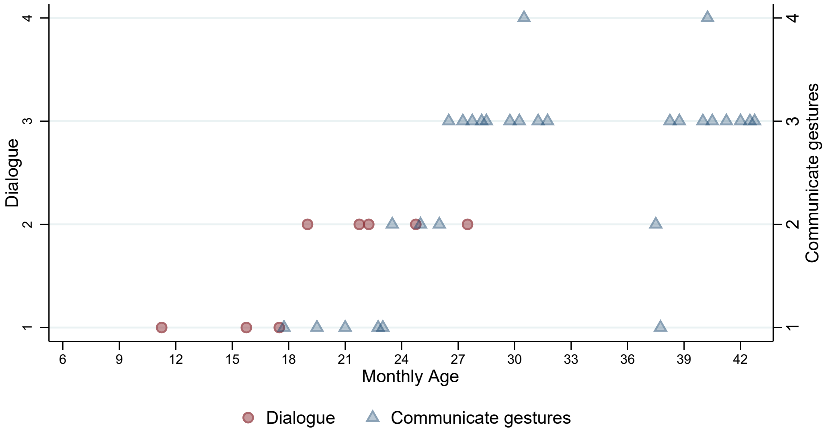
As the child grows, the caregiver progresses from simply narrating events to building on words the child has learned to scaffold language development.

C. Communicate gestures

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

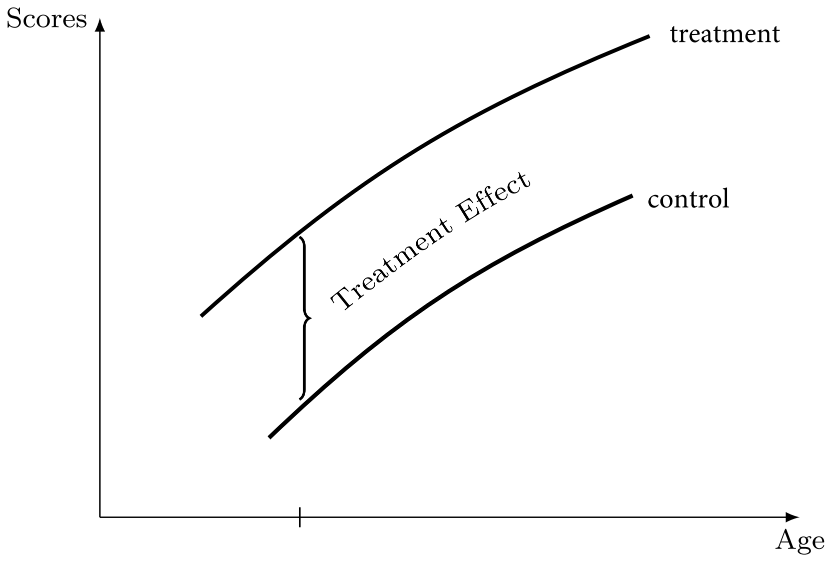


Note: Dialogue level 1: Talking to the child; level 2: Using the words the child learned to create conversation.
Communicate gestures level 1: the child listens to simple instructions; and level 4: the child can act as other roles, e.g., father, mother.

B Maturation and Exposure Effects

Maturation and exposure effects in learning contaminate tests of exchangeability, so we need to adjust the age-by-age probabilities of attaining mastery for maturation effects $\tau(a)$.

Figure B.1: Example of Maturation and Exposure Effects



Age Adjustment

We estimate maturation effects and exposure effects using the following logit model:

$$\Pr(D(s,\ell,a)) = \frac{\exp(\tau(a) + \kappa^1(s,\ell,a)|\text{treated})}{1+\exp(\tau(a) + \kappa^1(s,\ell,a)|\text{treated})}$$

where $D(s,\ell,a)$ is the indication of achieving the goal of the task of weekly age a, at difficulty level ℓ for skill type s.[6] Therefore the exposure effects of the intervention can be calculated by the following equation:

$$\ln \frac{\frac{Pr(D(s,\ell,a)=1|\text{treated})}{1-Pr(D(s,\ell,a)=1|\text{treated})}}{\frac{Pr(D(s,\ell,a)=1|\text{untreated})}{1-Pr(D(s,\ell,a)=1|\text{untreated})}} = \kappa^1(s,\ell,a) - \kappa^0(s,\ell,a)$$

Table B.1 presents the estimates of maturation effects $\tau(a)$ and the exposure effects $\kappa^1 - \kappa^0$. We find significant exposure effects for language, cognitive, and fine motor skill tasks. This finding provides another evidence that the intervention fosters the child's multiple skill development.

---

[6] "0" and "1" superscripts denote treatment status: "0" means the control group and "1" means the treatment group.

Table B.1: Treatment Effects Adjusting for Task Performance Maturation Effects for Each Skill (Log Odds Ratio)

	Language	Cognitive	Fine Motor	Gross Motor
Enrolled > 1 Month	0.722^{***}	0.804^{***}	1.347^{***}	0.253
$(\kappa^1 - \kappa^0)$	(0.196)	(0.298)	(0.360)	(0.367)
Monthly Age	0.104^{***}	0.155^{***}	0.147^{***}	-0.023
$(\tau(a))$	(0.029)	(0.020)	(0.035)	(0.032)
Monthly Age^2	-0.001^{***}	-0.002^{***}	-0.002^{***}	0.001
$(\tau(a))$	(0.001)	(0.000)	(0.001)	(0.001)
Enrolled > 1 Month × Monthly Age	-0.002	-0.005	-0.041^{*}	-0.004
	(0.009)	(0.012)	(0.016)	(0.018)
Male	-0.094^{***}	-0.099^{***}	-0.190^{***}	-0.147^{*}
	(0.084)	(0.079)	(0.122)	(0.100)
Constant	-1.873^{***}	-2.670^{***}	-1.721^{***}	0.938^{**}
	(0.414)	(0.360)	(0.545)	(0.403)

1. "Enrolled >1 Month" is the indicator variable for the children who enrolled longer than one month when home visits were implemented.

2. "Enrolled > 1 Month X Monthly Age" is the interaction term of the indicator variable (see footnote 1) and the monthly age of children.

3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses are clustered at village level.

C Measures of Interactions and Factor Model of Interaction

C.1 Measures of Interactions

China REACH collects weekly records of child performance on lessons that measure child development during the weekly home visit intervention. The supervisors record home visitor, parent, and child interaction activities at least once per month, making it possible for us to examine their impacts. These measures are only recorded for the treatment group. We exploit variation within the treatment group, which, as we document below, is substantial.

We have detailed measures to evaluate the interaction quality between home visitors and caregivers and that between home visitors and the visited children during home visits. These observation-based measures were recorded by the program supervisors who randomly visited each household at least once per month at randomly selected times. During the home visit, the program supervisor evaluated the home visit's quality in three dimensions: the quality of the home visitor's teaching ability, the interaction quality between home visitor and caregiver, and the interaction quality between the home visitor and the child. The measures we use for the interaction quality are listed in Table C.1, and the measures we use for teaching quality are listed in Table C.2.

Table C.1: Observed Interactions

Between Home Visitor and Caregiver
Has the home visitor explained the task content and lesson target to the caregiver?
Has the home visitor shown the lessons and given examples to the caregiver?
Does the home visitor ask the caregiver to play the lessons with the child alone?
Does the caregiver ask the home visitor about lessons in the next week?
Has the home visitor listened to the caregiver?
Has the home visitor answered the caregiver's questions?
Has the home visitor asked for the caregiver's opinions?
Does the home visitor encourage and help the caregiver?
Is the relationship between the home visitor and caregiver friendly, understandable, and cooperated?
Has the home visitor discussed with a caregiver or other persons about the content not related to the home visiting?
Between Home Visitor and Child
Has the home visitor shown the lessons and given examples to the child?
Has the home visitor explained the lesson to the child?
Does the home visitor listen to the child and respond to the child's voice or action?
Does the home visitor praise the child when the child tries to master one task?
Does the home visitor use language to communicate with the child when the child is completing the lessons?
Does the home visitor give the child enough time to explore the materials and finish the lessons?
Is the relationship between the home visitor, and the child-friendly, understandable, and cooperative?

Note: The interaction quality measures are recorded by the supervisor of the program at least once per month.

Table C.2: Home Visitor's Teaching Quality during Home Visiting

Does the home visitor bring the curriculum to the household?
Does the home visitor properly use the curriculum?
Has the home visitor prepared for the home visit in advance?
Has the home visitor chosen the teaching materials and tasks which are suitable to the child's age?
Does the home visitor focus on language development?

Note: The interaction quality measures are recorded by the supervisor of the program at least once per month.

C.2 Factor Model for Summarizing Interactions

As documented above, we have detailed measures on the home visitors' teaching ability and the interaction measures between the home visitor and caregiver (child). To summarize, we estimate the latent factors of home visitor's teaching ability and the interaction quality factors between home visitor and caregiver (child).

Denote $M_{ia}^{j,l}$ as the measure j at household i at the child's age a, and γ_{ia}^{l} is the latent factors l representing different factors (i.e., teaching ability, the interaction quality between home visitor and caregiver, and the interaction quality between the home visitor and child).

$$M_{ia}^{j,l} = X_{ia}'\beta + \alpha^{j}\gamma_{ia}^{l} + \epsilon_{ia}^{j,l} \tag{C.1}$$

Since the measure of M_{ia} is a categorical variable, we use an ordered probit model with latent factor γ and estimate the factor model by MLE assuming ϵ_{ia} is from normal distribution with zero mean.

We estimate the latent factor γ^{l} based using the Empirical Bayes method: the empirical conditional posterior distribution of the latent factor is given by

$$g(\gamma^l|M^l, X; \beta, \alpha) = \frac{\mu(M^l|X, \gamma^l; \beta, \alpha, \phi(\gamma^l))}{\int \mu(M^l|X, \gamma^l; \beta, \alpha, \phi(\gamma^l)) d\gamma^l} \tag{C.2}$$

Therefore, the latent factor estimate for latent factor l is given as $\hat{\gamma}^l = \int \gamma g(\gamma|M^l, X; \beta, \alpha) d\gamma^l$. The prior distribution of ϕ is based on estimated factor distribution. In Table C.3, we report the estimates of three factor distributions.

Table C.3: Prior Variances for Latent Factors

	Variance
Interaction between Home Visitor and Caregiver	0.685
	(0.046)
Interaction between Home Visitor and Child	2.914
	(0.200)
Teaching Ability	0.603
	(0.049)

Table C.4: Factor Model: Teacher Ability

	Measures Index				
	a6	a7	a8	a9	a10
Monthly Age	-0.008***	-0.034***	-0.021***	0.009***	-0.024***
	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)
Male	-0.148***	-0.043	-0.023	-0.274***	-0.213***
	(0.038)	(0.062)	(0.053)	(0.034)	(0.040)
Factor Loading	-1.000	-2.033***	-1.605***	-0.324***	-1.332***
	(0.000)	(0.147)	(0.102)	(0.034)	(0.072)
Cut 1	0.783**	1.217**	1.256**	1.221**	0.230**
	(0.068)	(0.118)	(0.098)	(0.061)	(0.072)
Cut 2	3.540**	4.867**	4.000**	3.512**	2.168**
	(0.139)	(0.278)	(0.176)	(0.147)	(0.087)
Cut 3		5.121**			4.023**
		(0.311)			(0.242)
Variance of the latent factor			0.603***		
(Teaching ability)			(0.049)		

1. Each variable represents a categorical variable evaluating Teacher's ability. Each variable corresponds to the following questions. a6: Does the home visitor bring the curriculum to the household? a7: Does the home visitor properly use the curriculum? a8 : Has the home visitor prepared the home visit in advance? a9: Has the home visitor chosen the teaching materials and tasks which are suitable? and a10 : Does the home visitor focus on language development?

2. All the measures are categorical variables with four categories. In ordered probit model, we have three cut off intercepts. The four categories are: (1) well done (2) basically achieve (3) not enough (4) not achieve at all.

3. Since the small values mean the higher quality in all the measures, we normalize the loading of the first measure to -1, which makes the larger latent factor values mean better quality.

4. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.5: Factor Model: Interactions Quality between Home Visitor and Caregiver

	Measures Index									
	a18	a11	a12	a15	a16	a17	a19	a20	a21	a22
Monthly Age	0.013***	-0.014***	-0.023***	-0.024***	-0.014***	-0.009***	-0.032***	-0.002	-0.018***	-0.003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Male	-0.253***	-0.138***	-0.017	-0.211***	-0.123***	-0.022	-0.368***	0.081**	-0.174***	-0.359***
	(0.038)	(0.041)	(0.049)	(0.041)	(0.037)	(0.030)	(0.057)	(0.033)	(0.048)	(0.059)
Factor Loading	-1.000	-1.073***	-1.489***	-1.268***	-1.008***	-0.781***	-1.791***	-0.944***	-1.521***	-1.896***
	(0.000)	(0.049)	(0.066)	(0.054)	(0.045)	(0.036)	(0.081)	(0.041)	(0.067)	(0.087)
Cut 1	1.289**	0.864**	0.979**	0.325**	0.582**	0.168**	0.959**	0.646**	0.872**	1.745**
	(0.069)	(0.073)	(0.088)	(0.073)	(0.066)	(0.055)	(0.100)	(0.060)	(0.085)	(0.110)
Cut 2	3.294**	2.577**	3.622**	2.080**	2.068**	1.616**	2.917**	1.773**	3.277**	4.983**
	(0.090)	(0.087)	(0.131)	(0.081)	(0.074)	(0.059)	(0.121)	(0.063)	(0.113)	(0.181)
Cut 3	4.486**	4.267**	4.678**	2.237**	2.959**	3.680**	4.390**	3.182**	4.611**	5.459**
	(0.163)	(0.293)	(0.248)	(0.082)	(0.093)	(0.155)	(0.189)	(0.085)	(0.222)	(0.221)
Variance of the latent factor	0.685***									
(Interaction: Home Visitor and Caregiver)	(0.046)									

1. The variables represent a categorical variable evaluating interaction quality. Each variable corresponds to the following questions:
a18: Has the home visitor listened to the caregiver? a11 : Has the home visitor explained the task content and task target to the caregiver?
a12: Has the home visitor shown the tasks and given an example to the caregiver? a15 : Does the home visitor ask the caregiver to play the tasks with the child alone?
a16: Does the caregiver answer the home visitor about what will play in the next week? a17: Has the home visitor discuss with a caregiver or other persons about the content?
a19: Has the home visitor answered caregiver's question? a20: Has the home visitor asked for the caregiver's opinions? a21: Does the home visitor encourage and help the caregiver? a22: Is the relationship between the home visitor and caregiver friendly?

2. All the measures are categorical variables with four categories. In ordered probit model, we have three cut off intercepts. The four categories are:
(1) well done (2) basically achieve (3) not enough (4) not achieved at all.

3. Since the small values mean the higher quality in all the measures, we normalize the loading of the first measure to -1, which makes the larger latent factor values mean better quality.

4. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.6: Factor Model: Interactions Quality between Home Visitor and Child

	Measures Index						
	a23	a13	a14	a24	a25	a26	a27
Monthly Age	-0.021***	-0.019***	0.006**	-0.021***	0.027***	-0.005	-0.008**
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Male	-0.162***	-0.149***	-0.115***	-0.017	-0.205***	-0.149***	-0.255***
	(0.056)	(0.045)	(0.040)	(0.056)	(0.050)	(0.055)	(0.058)
Factor Loading:	-1.000	-0.685***	-0.664***	-1.099***	-0.918***	-1.053***	-0.992***
	(0.000)	(0.029)	(0.028)	(0.048)	(0.042)	(0.048)	(0.044)
Cut 1	1.249**	0.868**	1.105**	1.063**	2.130**	1.502**	1.743**
	(0.102)	(0.081)	(0.074)	(0.103)	(0.098)	(0.105)	(0.109)
Cut 2	3.061**	2.781**	3.184**	2.958**	3.686**	3.713**	3.731**
	(0.123)	(0.101)	(0.091)	(0.119)	(0.114)	(0.134)	(0.143)
Cut 3	3.570**	3.029**	3.550**	3.830**	4.575**	4.139**	3.856**
	(0.140)	(0.110)	(0.101)	(0.146)	(0.136)	(0.150)	(0.149)
Variance of the latent factor				2.914***			
(Interaction: Home Visitor and Child)				(0.200)			

1.Each variable represents a categorical variable evaluating interaction quality. Each variable corresponds to the following questions: a23 : Does the home visitor listen to the child and respond to the child's voice? a13: Has the home visitor shown the tasks and given an example to the child? a14: Has the home visitor explained the task to the child? a24: Does the home visitor praise the child when the child tries to finish one task? a25: Does the home visitor use language to communicate with the child when the child? a26: Does the home visitor give the child enough time to explore the materials? a27: Is the relationship between the home visitor and the child is friendly?

2. All the measures are categorical variables with four categories. In ordered probit model, we have three cut off intercepts. The four categories are: (1) well done (2) basically achieved (3) not enough (4) not achieved at all.

3. Since the small values mean the higher quality in all the measures, we normalize the loading of the first measure to -1, which makes the larger latent factor values mean better quality.

4. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D The Causal Effects of Interaction Quality on Learning Measures

This appendix reports the impact of interactions on the various measures of knowledge of skill at different levels.

Controlling for Endogeneity of Interactions

Interaction measures are likely correlated with person-specific factors. If latent skill factors change with age, the traditional fixed-effect method does not work. Instead, we use instrumental variables formed in the following manner. Each week the home visitors visit multiple households, some with children who live in the same village as the child, and some with children who live in another village. We use the interaction quality measure for the same home visitor with other household caregivers and children as instrumental variables for the interaction quality between the home visitor and caregiver and child for a given household in a village. We use the mean (maximum and minimum) of the home visitor interaction quality measures with caregivers and children at all other households who are not in the same village with the child.

We find substantial effects of interactions on Denver scores that are especially strong when the measures are instrumented.

D.1 Time to Mastery

Table D.1: The Effects of Interactions on the Time to Mastery of Language Tasks at Each Level (OLS)

	Language Task Difficulty Levels									
	≤2	3	4	5	6	7	8	9	10	11
Interaction Quality:	-0.233	-0.322*	-0.131	-0.287	-0.510**	-0.468	-0.301	-0.167*	-0.225	0.021
Home Visitor and Caregiver	(0.530)	(0.167)	(0.206)	(0.310)	(0.228)	(0.349)	(0.271)	(0.088)	(0.154)	(0.072)
Interaction Quality:	-0.163	0.017	-0.130**	-0.112	-0.032	-0.335**	-0.054	-0.004	0.039	-0.114***
Home Visitor and Child	(0.144)	(0.041)	(0.049)	(0.115)	(0.096)	(0.127)	(0.065)	(0.034)	(0.027)	(0.042)
Teaching Ability	0.103	0.021	-0.051	0.163	0.176	0.515	-0.452	-0.104	0.160**	-0.248*
	(0.293)	(0.171)	(0.177)	(0.318)	(0.164)	(0.328)	(0.301)	(0.107)	(0.076)	(0.141)
Grandmother Rearing[1]	-0.263	0.134	-0.108	0.412	-0.015	0.424	0.394**	0.217**	0.339***	0.156*
	(0.343)	(0.146)	(0.126)	(0.454)	(0.143)	(0.258)	(0.178)	(0.099)	(0.119)	(0.087)
Monthly Age	-0.120***	-0.074***	-0.096***	-0.132***	-0.085***	-0.041**	-0.015	-0.020**	-0.014	-0.035***
	(0.024)	(0.021)	(0.018)	(0.038)	(0.016)	(0.016)	(0.014)	(0.010)	(0.009)	(0.010)
Constant	2.293***	2.563***	3.604***	3.189***	2.704***	1.929*	2.070**	1.911***	1.481***	2.003***
	(0.826)	(0.568)	(0.413)	(0.729)	(0.741)	(1.126)	(0.797)	(0.379)	(0.509)	(0.414)

1. % of home visits when grandmother is the primary caregiver.

2. The estimates reported in the table do not include instrumental variables.

3. The variables of teaching ability, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.

4. Time to Mastery is defined as the number of tasks a child takes until the first success (inclusive) at each difficulty level by each skill type.

5. Standard errors are reported in parentheses and clustered at village level.

6.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.2: The Effects of Interactions on the Time to Mastery of Language Tasks at Each Level (IV)

	Language Task Difficulty Levels									
	≤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)
Cragg-Donald F	15.414	13.948	16.736	8.400	40.491	55.196	50.644	43.970	59.752	44.872
Kleibergen-Paap LM	25.461	16.525	50.765	8.837	51.326	32.711	136.875	56.539	112.133	39.504
Hansen J	2.985	1.191	8.313	0.562	1.362	1.860	4.479	0.021	0.013	0.968
P-Value(Hansen J)	0.225	0.551	0.016	0.755	0.506	0.395	0.106	0.989	0.993	0.616

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
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 at the previous difficulty level until the first success (inclusive) at each difficulty level by each skill type.
6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics. For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.
7. Standard errors are reported in parentheses and clustered at village level.
8.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: The Effects of Interactions on the Time to Mastery of Cognitive Tasks at Each Level (OLS)

	Cognitive Task Difficulty Levels										
	≤2	3	4	5	6	7	8	9	10	11	12
Interaction Quality: Home Visitor and Caregiver	-0.305 (0.186)	-0.166** (0.069)	-0.031 (0.052)	-0.422** (0.203)	-0.214 (0.170)	-0.184 (0.246)	0.011 (0.129)	-0.155*** (0.057)	-0.214 (0.140)	-0.132 (0.172)	-0.200** (0.085)
Interaction Quality: Home Visitor and Child	-0.246** (0.096)	-0.035 (0.028)	-0.024 (0.023)	-0.158 (0.110)	-0.065 (0.065)	-0.057 (0.065)	-0.050 (0.039)	0.016 (0.013)	-0.059* (0.031)	0.015 (0.057)	0.006 (0.041)
Teaching Ability	0.049 (0.196)	0.082 (0.067)	0.043 (0.080)	-0.212 (0.308)	-0.249 (0.186)	-0.167 (0.250)	-0.159 (0.140)	0.052 (0.047)	0.229* (0.124)	-0.180 (0.184)	-0.220** (0.101)
Grandmother Rearing[1]	-0.001 (0.167)	0.020 (0.067)	0.022 (0.059)	0.148 (0.229)	0.366** (0.162)	0.279 (0.203)	0.148 (0.133)	0.063 (0.042)	0.248* (0.139)	0.674*** (0.197)	0.118 (0.101)
Monthly Age	-0.068*** (0.012)	-0.002 (0.006)	0.000 (0.008)	-0.022 (0.014)	-0.035** (0.013)	-0.010 (0.016)	-0.009 (0.012)	-0.010** (0.004)	-0.000 (0.008)	0.047 (0.030)	-0.003 (0.006)
Constant	2.817*** (0.391)	1.272*** (0.202)	1.088*** (0.223)	2.398*** (0.609)	2.662*** (0.469)	2.679*** (0.766)	0.759 (0.520)	1.252*** (0.139)	1.462*** (0.498)	-1.216 (1.130)	0.911*** (0.309)

1. % of home visits when grandmother is the primary caregiver.
2. The estimates reported in the table do not include instrumental variables.
3. The variables of teaching ability, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
4. Time to Mastery is defined as the number of tasks a child takes until the first success (inclusive) at each difficulty level by each skill type.
5. Standard errors are reported in parentheses and clustered at village level.
6.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4: The Effects of Interactions on the Time to Mastery of Cognitive Tasks at Each Level (IV)

	Cognitive Task Difficulty Levels										
	≤2	3	4	5	6	7	8	9	10	11	12
Interaction Quality Home Visitor and Caregiver	-0.923* (0.515)	-0.212** (0.108)	0.007 (0.092)	-0.819*** (0.190)	-0.710*** (0.276)	-0.699* (0.366)	-0.259* (0.150)	-0.208*** (0.060)	-0.669*** (0.215)	-0.466** (0.189)	-0.196 (0.131)
Interaction Quality Home Visitor and Child	-0.082 (0.130)	-0.003 (0.028)	-0.052** (0.021)	-0.091 (0.077)	-0.042 (0.068)	0.050 (0.100)	-0.015 (0.055)	0.019* (0.010)	0.004 (0.042)	0.015 (0.050)	-0.049 (0.045)
Teaching Ability	0.402 (0.548)	0.261** (0.101)	-0.245** (0.123)	0.770** (0.342)	0.600 (0.370)	-0.345 (0.434)	0.231 (0.204)	0.054 (0.060)	0.503*** (0.167)	0.177 (0.274)	-0.205 (0.164)
Grandmother Rearing[1]	0.032 (0.255)	-0.002 (0.062)	-0.027 (0.071)	0.437* (0.258)	0.436** (0.176)	0.006 (0.225)	0.283* (0.155)	0.100** (0.045)	0.336** (0.162)	0.793*** (0.209)	0.117 (0.088)
Monthly Age	-0.057*** (0.012)	-0.007 (0.004)	0.007 (0.010)	-0.012 (0.012)	-0.018 (0.021)	0.025 (0.026)	0.032* (0.018)	-0.001 (0.003)	0.002 (0.009)	0.067** (0.032)	-0.011 (0.011)
Constant	2.309*** (0.762)	1.033*** (0.259)	0.970*** (0.311)	1.726*** (0.540)	2.463*** (0.687)	2.666* (1.541)	0.516 (0.498)	1.025*** (0.095)	1.222** (0.532)	-1.019 (0.884)	1.601*** (0.453)
Cragg-Donald F	43.494	34.803	22.807	43.648	48.213	96.371	49.372	36.137	54.441	34.974	17.043
Kleibergen-Paap LM	65.949	62.963	43.384	53.898	55.824	89.574	72.079	54.228	90.675	34.408	52.252
Hansen J	1.962	5.604	2.901	0.858	0.779	3.639	3.913	0.754	2.669	1.392	2.451
P-Value(Hansen J)	0.375	0.061	0.234	0.651	0.678	0.162	0.141	0.686	0.263	0.499	0.294

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
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 at the previous difficulty level until the first success (inclusive) at each difficulty level by each skill type.
6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics. For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.
7. Standard errors are reported in parentheses and clustered at village level.
8.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5: The Effects of Interactions on the Time to Mastery of Fine Motor Tasks at Each Level (OLS)

	Fine Motor Task Difficulty Levels						
	1	2	3	4	5	6	7
Interaction Quality: Home Visitor and Caregiver	-0.032	-0.024	-0.190**	-0.151	-0.086*	-0.170*	-0.108*
	(0.102)	(0.038)	(0.081)	(0.092)	(0.044)	(0.100)	(0.063)
Interaction Quality: Home Visitor and Child	-0.056	-0.012	0.036	-0.067	0.010	0.018	-0.020
	(0.043)	(0.017)	(0.041)	(0.058)	(0.023)	(0.041)	(0.035)
Teaching Ability	0.034	0.018	-0.136	0.175*	0.022	0.181	0.105*
	(0.114)	(0.045)	(0.140)	(0.100)	(0.079)	(0.113)	(0.057)
Grandmother Rearing[1]	-0.171	-0.048	0.042	0.233*	-0.004	0.322**	0.035
	(0.131)	(0.033)	(0.074)	(0.131)	(0.042)	(0.143)	(0.061)
Monthly Age	-0.054***	-0.008***	-0.002	-0.002	0.001	0.003	0.005
	(0.013)	(0.003)	(0.013)	(0.008)	(0.004)	(0.013)	(0.004)
Constant	2.414***	1.209***	1.787***	1.121***	0.913***	0.482	0.928***
	(0.521)	(0.091)	(0.443)	(0.295)	(0.152)	(0.646)	(0.215)

1. % of home visits when grandmother is the primary caregiver.

2. The estimates reported in the table do not include instrumental variables.

3. The variables of teaching ability, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.

4. Time to Mastery is defined as the number of tasks a child takes until the first success (inclusive) at each difficulty level by each skill type.

5. Standard errors are reported in parentheses and clustered at village level.

6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.6: The Effects of Interactions on the Time to Mastery of Fine Motor Tasks at Each Level (IV)

	Fine Motor Task Difficulty Levels						
	1	2	3	4	5	6	7
Interaction Quality	-0.276	-0.100	-0.374**	-0.286**	-0.183***	-0.433***	-0.087
Home Visitor and Caregiver	(0.191)	(0.079)	(0.168)	(0.114)	(0.043)	(0.092)	(0.100)
Interaction Quality	-0.024	0.014	0.094**	0.039	0.036***	0.048*	-0.035
Home Visitor and Child	(0.080)	(0.018)	(0.046)	(0.026)	(0.011)	(0.029)	(0.046)
Teaching Ability	-0.215	0.057	-0.192	0.186	0.115*	0.367***	0.034
	(0.276)	(0.095)	(0.255)	(0.173)	(0.066)	(0.108)	(0.087)
Grandmother Rearing[1]	-0.105	-0.015	-0.072	0.270	0.007	0.228*	0.167*
	(0.142)	(0.053)	(0.133)	(0.184)	(0.048)	(0.138)	(0.090)
Monthly Age	-0.050***	-0.009***	0.015	-0.005	0.001	0.015	0.008***
	(0.014)	(0.002)	(0.017)	(0.008)	(0.004)	(0.012)	(0.003)
Constant	3.081***	1.183***	1.333**	1.332***	1.058***	0.524	0.631***
	(0.880)	(0.168)	(0.656)	(0.333)	(0.222)	(0.557)	(0.187)
Cragg-Donald F	22.321	29.276	52.586	38.079	37.828	33.491	18.600
Kleibergen-Paap LM	30.921	44.915	53.127	59.348	33.557	58.955	51.656
Hansen J	5.685	5.263	2.348	3.647	2.394	0.023	1.063
P-Value(Hansen J)	0.058	0.072	0.309	0.161	0.302	0.989	0.588

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.

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 at the previous difficulty level until the first success (inclusive) at each difficulty level by each skill type.

6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics.
For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.

7. Standard errors are reported in parentheses and clustered at village level.

8.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.7: The Effects of Interactions on the Time to Mastery of Gross Motor Tasks at Each Level (OLS)

	Gross Motor Task Difficulty Levels						
	≤ 3	4	5	6	7	8	9
Interaction Quality:	0.001	-0.157*	-0.041*	-0.059*	-0.032	-0.058	-0.017
Home Visitor and Caregiver	(0.032)	(0.078)	(0.022)	(0.030)	(0.043)	(0.066)	(0.033)
Interaction Quality:	0.007	-0.046	0.017*	-0.007	-0.010	-0.004	-0.013
Home Visitor and Child	(0.006)	(0.034)	(0.010)	(0.006)	(0.026)	(0.031)	(0.011)
Teaching Ability	-0.017	-0.010	0.018	0.008	-0.108**	0.024	0.060
	(0.020)	(0.083)	(0.023)	(0.044)	(0.050)	(0.076)	(0.040)
Grandmother Rearing[1]	-0.012	-0.055	-0.030**	0.076*	-0.008	0.032	-0.028
	(0.010)	(0.082)	(0.014)	(0.044)	(0.039)	(0.061)	(0.038)
Monthly Age	-0.003	-0.059***	-0.001	-0.009***	-0.004	-0.005	0.000
	(0.002)	(0.014)	(0.001)	(0.002)	(0.003)	(0.007)	(0.004)
Constant	1.032***	2.484***	1.012***	1.192***	1.454***	1.704***	1.037***
	(0.037)	(0.423)	(0.031)	(0.078)	(0.143)	(0.311)	(0.134)

1. % of home visits when grandmother is the primary caregiver.

2. The estimates reported in the table do not include instrumental variables.

3. The variables of teaching ability, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.

4. Time to Mastery is defined as the number of tasks a child takes until the first success (inclusive) at each difficulty level by each skill type.

5. Standard errors are reported in parentheses and clustered at village level.

6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.8: The Effects of Interactions on the Time to Mastery of Gross Motor Tasks at Each Level (IV)

	Gross Motor Task Difficulty Levels						
	≤ 3	4	5	6	7	8	9
Interaction Quality Home Visitor and Caregiver	0.011 (0.012)	-0.351*** (0.113)	-0.000 (0.000)	0.015 (0.046)	-0.002 (0.052)	-0.158 (0.105)	-0.013 (0.028)
Interaction Quality Home Visitor and Child	0.000 (0.001)	0.010 (0.044)	0.000 (0.000)	-0.008 (0.005)	-0.021 (0.017)	0.020 (0.028)	-0.004 (0.010)
Teaching Ability	-0.010 (0.011)	0.104 (0.201)	0.001 (0.001)	-0.032 (0.079)	0.009 (0.055)	0.122 (0.126)	0.100* (0.055)
Grandmother Rearing[1]	-0.004 (0.005)	0.064 (0.115)	-0.000 (0.000)	0.073 (0.065)	0.008 (0.046)	0.097 (0.076)	-0.001 (0.040)
Monthly Age	0.000 (0.000)	-0.064*** (0.019)	-0.000 (0.000)	-0.006*** (0.002)	-0.004*** (0.001)	0.001 (0.007)	0.004 (0.005)
Constant	0.996*** (0.006)	1.977*** (0.381)	1.001*** (0.001)	1.029*** (0.057)	1.238*** (0.133)	1.086*** (0.271)	0.810*** (0.153)
Cragg-Donald F	6.009	12.714	20.248	22.593	46.363	57.358	57.265
Kleibergen-Paap LM	6.554	18.880	59.632	24.859	93.066	97.683	53.691
Hansen J	0.500	3.784	1.054	2.940	2.079	0.088	2.080
P-Value(Hansen J)	0.779	0.151	0.590	0.230	0.354	0.957	0.354

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.

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 at the previous difficulty level until the first success (inclusive) at each difficulty level by each skill type.

6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics.
For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.

7. Standard errors are reported in parentheses and clustered at village level.

8.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Instability

Table D.9: The Effects of Interactions on the Instability of Language Tasks at Each Level (OLS)

	Language Task Difficulty Levels									
	2	3	4	5	6	7	8	9	10	11
Interaction Quality: Home Visitor and Caregiver	-0.058	0.026	-0.107	0.070	-0.012	-0.051	-0.078*	-0.080	-0.157***	-0.061*
	(0.075)	(0.056)	(0.068)	(0.067)	(0.044)	(0.042)	(0.044)	(0.053)	(0.043)	(0.031)
Interaction Quality: Home Visitor and Child	0.004	-0.027	-0.016	-0.002	-0.024	-0.011	-0.024**	-0.006	0.012	-0.004
	(0.024)	(0.030)	(0.024)	(0.045)	(0.015)	(0.011)	(0.011)	(0.016)	(0.022)	(0.016)
Teaching Ability	-0.031	-0.090	-0.065	0.091	-0.037	-0.054	0.026	-0.070	-0.015	-0.016
	(0.076)	(0.078)	(0.076)	(0.120)	(0.044)	(0.048)	(0.049)	(0.042)	(0.066)	(0.044)
Grandmother Rearing[1]	-0.054	0.001	0.120	0.064	-0.028	0.086**	0.079**	0.034	0.037	0.148***
	(0.060)	(0.075)	(0.087)	(0.156)	(0.058)	(0.036)	(0.031)	(0.053)	(0.044)	(0.053)
Monthly Age	0.002	0.004	-0.002	0.008	-0.002	-0.003	-0.001	0.003	-0.002	-0.008**
	(0.007)	(0.009)	(0.007)	(0.013)	(0.004)	(0.002)	(0.003)	(0.005)	(0.004)	(0.004)
Constant	0.175	0.317	0.114	0.486	0.094	0.118	0.043	0.099	-0.092	0.095
	(0.271)	(0.243)	(0.313)	(0.311)	(0.186)	(0.173)	(0.106)	(0.140)	(0.187)	(0.179)

1. % of home visits when grandmother is the primary caregiver.
2. The estimates reported in the table do not include instrumental variables.
3. The variables of teaching ability, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
4. For intervention tasks, instability is defined as fraction of fails after the first success at each difficulty level by each skill type.
5. Standard errors are reported in parentheses and clustered at village level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.10: The Effects on the Instability of Language Tasks at Each Level (IV)

	Language Task Difficulty Levels									
	2	3	4	5	6	7	8	9	10	11
Interaction Quality Home Visitor and Caregiver	-0.201 (0.148)	0.130 (0.105)	-0.227 (0.170)	0.149 (0.098)	-0.063 (0.075)	-0.155** (0.057)	-0.169** (0.062)	-0.082 (0.059)	-0.284*** (0.034)	-0.139*** (0.039)
Interaction Quality Home Visitor and Child	-0.001 (0.046)	-0.033 (0.050)	0.024 (0.031)	-0.006 (0.047)	-0.021 (0.016)	-0.018 (0.017)	-0.004 (0.010)	-0.015 (0.023)	0.057*** (0.012)	-0.016 (0.017)
Teaching Ability	0.240 (0.172)	-0.404** (0.156)	0.114 (0.185)	0.064 (0.163)	-0.010 (0.088)	0.012 (0.071)	0.112* (0.051)	-0.181* (0.079)	-0.022 (0.053)	0.049 (0.066)
Grandmother Rearing	-0.036 (0.075)	-0.099 (0.079)	0.070 (0.111)	0.230 (0.133)	-0.034 (0.076)	0.057 (0.043)	0.060 (0.036)	0.034 (0.063)	-0.031 (0.048)	0.111* (0.055)
Monthly Age	0.005 (0.010)	-0.003 (0.009)	-0.010 (0.008)	0.023 (0.016)	0.002 (0.004)	-0.001 (0.002)	0.002 (0.004)	0.003 (0.004)	-0.005 (0.003)	-0.011** (0.004)
Constant	0.267 (0.311)	0.801** (0.245)	0.571 (0.383)	0.266 (0.327)	0.595** (0.190)	0.359* (0.142)	0.093 (0.149)	0.375** (0.144)	0.178 (0.131)	0.448** (0.147)
Cragg-Donald F	10.927	11.340	11.424	4.476	32.708	55.509	48.269	36.623	36.041	38.171
Kleibergen-Paap LM	17.683	10.518	27.827	6.515	39.072	33.716	130.377	41.069	63.937	36.011
Hansen J	2.513	3.479	6.998	0.547	8.562	2.129	2.284	3.196	7.400	0.053
P-Value(Hansen J)	0.285	0.176	0.030	0.761	0.014	0.345	0.319	0.202	0.025	0.974

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.

4. The instrumental variables include mean, max, and min of other village interaction measures through the same visitor.

5. Instability is defined as fraction of fails after the first success at each difficulty level by each skill type.

6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics. For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.

7. Standard errors are reported in parentheses and clustered at village level.

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

Table D.11: The Effects of Interactions on the Instability of Cognitive Tasks at Each Level (OLS)

	Cognitive Task Difficulty Levels									
	3	4	5	6	7	8	9	10	11	12
Interaction Quality: Home Visitor and Caregiver	-0.055	-0.158**	-0.120**	-0.055*	-0.020	-0.036	0.011	-0.023	-0.136**	-0.165***
	(0.062)	(0.078)	(0.048)	(0.031)	(0.034)	(0.048)	(0.085)	(0.051)	(0.062)	(0.057)
Interaction Quality: Home Visitor and Child	-0.018	-0.046*	-0.017	-0.010	-0.026*	-0.023	-0.039*	-0.015	-0.016	0.003
	(0.038)	(0.025)	(0.018)	(0.014)	(0.015)	(0.016)	(0.023)	(0.026)	(0.034)	(0.023)
Teaching Ability	0.028	0.011	-0.011	0.015	-0.049	0.003	0.123	-0.111*	-0.080	-0.013
	(0.081)	(0.113)	(0.045)	(0.027)	(0.030)	(0.035)	(0.109)	(0.057)	(0.064)	(0.074)
Grandmother Rearing[1]	0.228**	-0.106	0.011	0.019	0.052	0.069**	0.226**	0.001	0.090	0.002
	(0.096)	(0.068)	(0.043)	(0.021)	(0.037)	(0.032)	(0.091)	(0.035)	(0.060)	(0.067)
Monthly Age	-0.002	-0.004	0.000	-0.005*	-0.006*	-0.008***	-0.002	-0.003	-0.000	-0.000
	(0.019)	(0.024)	(0.003)	(0.003)	(0.003)	(0.003)	(0.011)	(0.006)	(0.006)	(0.011)
Constant	0.335	0.031	0.261**	0.240***	0.147	0.209**	-0.042	0.127	-0.022	0.003
	(0.474)	(0.550)	(0.127)	(0.085)	(0.130)	(0.102)	(0.402)	(0.222)	(0.260)	(0.445)

1. % of home visits when grandmother is the primary caregiver.
2. The estimates reported in the table do not include instrumental variables.
3. The variables of teaching ability, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
4. For intervention tasks, instability is defined as fraction of fails after the first success at each difficulty level by each skill type.
5. Standard errors are reported in parentheses and clustered at village level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.12: The Effects of Interactions on the Instability of Cognitive Tasks at Each Level (IV)

	Cognitive Task Difficulty Levels									
	3	4	5	6	7	8	9	10	11	12
Interaction Quality:	-0.021	-0.054	-0.226***	-0.105***	-0.075*	-0.050	-0.090	-0.082	-0.197***	-0.290***
Home Visitor and Caregiver	(0.051)	(0.049)	(0.065)	(0.032)	(0.043)	(0.060)	(0.105)	(0.088)	(0.065)	(0.069)
Interaction Quality:	-0.011	-0.028	0.012	0.006	-0.007	-0.023	-0.062***	-0.020	-0.002	0.022
Home Visitor and Child	(0.031)	(0.019)	(0.016)	(0.013)	(0.017)	(0.020)	(0.016)	(0.039)	(0.030)	(0.020)
Teaching Ability	0.001	-0.046	0.154**	0.082*	-0.075	-0.077	0.042	-0.056	-0.095	-0.053
	(0.099)	(0.111)	(0.067)	(0.045)	(0.061)	(0.059)	(0.142)	(0.073)	(0.076)	(0.066)
Grandmother Rearing[1]	0.136	-0.106**	-0.033	0.056**	0.024	0.022	0.151	0.012	0.082	0.016
	(0.100)	(0.052)	(0.036)	(0.023)	(0.034)	(0.034)	(0.103)	(0.034)	(0.061)	(0.087)
Monthly Age	0.004	0.030**	0.003	-0.002	-0.003	-0.008**	0.006	-0.003	0.000	0.001
	(0.019)	(0.015)	(0.003)	(0.002)	(0.003)	(0.003)	(0.012)	(0.005)	(0.007)	(0.013)
Constant	0.089	-0.527*	0.137*	0.149**	0.277***	0.373***	-0.038	0.260	0.204	0.137
	(0.395)	(0.316)	(0.077)	(0.062)	(0.084)	(0.092)	(0.409)	(0.172)	(0.288)	(0.528)
Cragg-Donald F	61.797	17.477	43.296	44.316	92.538	48.467	17.534	36.241	31.749	19.400
Kleibergen-Paap LM	57.084	20.919	53.824	66.880	76.859	58.993	17.665	32.570	47.164	50.080
Hansen J	5.203	0.891	3.856	4.889	2.807	3.318	3.811	0.333	1.759	4.927
P-Value(Hansen J)	0.074	0.640	0.145	0.087	0.246	0.190	0.149	0.847	0.415	0.085

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.

4. The instrumental variables include mean, max, and min of other village interaction measures through the same visitor.

5. For intervention tasks, instability is defined as fraction of fails after the first success at each difficulty level by each skill type.

6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics.
For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.

7. Standard errors are reported in parentheses and clustered at village level.

8.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.13: The Effects of Interactions on the Instability of Fine Motor Tasks at Each Level (OLS)

	Fine Motor Task Difficulty Levels					
	2	3	4	5	6	7
Interaction Quality: Home Visitor and Caregiver	-0.021	-0.150***	-0.041**	-0.098*	-0.188***	-0.145*
	(0.046)	(0.038)	(0.016)	(0.054)	(0.061)	(0.080)
Interaction Quality: Home Visitor and Child	-0.048*	-0.000	-0.013*	0.024	0.008	-0.037
	(0.026)	(0.016)	(0.007)	(0.019)	(0.023)	(0.034)
Teaching Ability	-0.147**	0.059	0.032	0.019	0.003	-0.052
	(0.067)	(0.042)	(0.026)	(0.079)	(0.066)	(0.103)
Grandmother Rearing[1]	-0.031	-0.008	0.021	0.044	0.124**	0.001
	(0.044)	(0.034)	(0.030)	(0.084)	(0.061)	(0.079)
Monthly Age	-0.008*	-0.003	0.003	0.012*	-0.008	-0.010
	(0.004)	(0.004)	(0.003)	(0.006)	(0.006)	(0.010)
Constant	0.367**	0.420**	-0.060	-0.523	0.294	0.522
	(0.144)	(0.166)	(0.100)	(0.336)	(0.229)	(0.605)

1. % of home visits when grandmother is the primary caregiver.
2. The estimates reported in the table do not include instrumental variables.
3. The variables of teaching ability, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
4. For intervention tasks, instability is defined as fraction of fails after the first success at each difficulty level by each skill type.
5. Standard errors are reported in parentheses and clustered at village level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.14: The Effects of Interactions on the Instability of Fine Motor Tasks at Each Level (IV)

	Fine Motor Task Difficulty Levels					
	2	3	4	5	6	7
Interaction Quality:	0.025	-0.166**	-0.022	-0.186***	-0.329***	-0.355***
Home Visitor and Caregiver	(0.064)	(0.077)	(0.022)	(0.060)	(0.052)	(0.053)
Interaction Quality:	-0.091***	0.011	-0.012	0.002	0.035*	0.017
Home Visitor and Child	(0.034)	(0.017)	(0.007)	(0.022)	(0.020)	(0.017)
Teaching Ability	-0.095	0.116	0.023	0.192**	0.175***	-0.105
	(0.101)	(0.083)	(0.046)	(0.094)	(0.061)	(0.150)
Grandmother Rearing^{1}	-0.015	-0.019	0.040	-0.008	0.067	-0.035
	(0.056)	(0.034)	(0.035)	(0.082)	(0.046)	(0.071)
Monthly Age	-0.011***	0.001	0.005*	0.010*	-0.004	0.010*
	(0.004)	(0.004)	(0.003)	(0.005)	(0.006)	(0.006)
Constant	0.395***	0.146	-0.088	-0.124	0.275	-0.146
	(0.103)	(0.097)	(0.090)	(0.186)	(0.228)	(0.233)
Cragg-Donald F	20.153	53.293	34.758	26.768	23.502	12.605
Kleibergen-Paap LM	45.091	56.010	53.790	24.785	41.115	28.433
Hansen J	3.033	0.581	5.281	2.322	1.025	1.220
P-Value(Hansen J)	0.220	0.748	0.071	0.313	0.599	0.543

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
4. The instrumental variables include mean, max, and min of other village interaction measures through the same visitor.
5. For intervention tasks, instability is defined as fraction of fails after the first success at each difficulty level by each skill type.
6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics. For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.
7. Standard errors are reported in parentheses and clustered at village level.
8.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.15: The Effects of Interactions on the Instability of Gross Motor Tasks at Each Level (OLS)

	Gross Motor Task Difficulty Levels				
	5	6	7	8	9
Interaction Quality: Home Visitor and Caregiver	-0.084	0.104	-0.036	-0.130*	-0.040
	(0.100)	(0.115)	(0.035)	(0.067)	(0.043)
Interaction Quality: Home Visitor and Child	-0.005	-0.046	-0.006	0.006	0.006
	(0.048)	(0.041)	(0.013)	(0.028)	(0.023)
Teaching Ability	0.098	-0.055	-0.083	0.001	0.032
	(0.113)	(0.121)	(0.054)	(0.055)	(0.072)
Grandmother Rearing[1]	-0.094	0.049	0.042	0.015	-0.039
	(0.136)	(0.116)	(0.036)	(0.064)	(0.042)
Monthly Age	0.058**	0.013	-0.007***	-0.009	-0.006**
	(0.025)	(0.016)	(0.002)	(0.007)	(0.003)
Constant	-0.402	-0.146	0.083	0.605***	0.271**
	(0.321)	(0.461)	(0.096)	(0.193)	(0.129)

1. % of home visits when grandmother is the primary caregiver.
2. The estimates reported in the table do not include instrumental variables.
3. The variables of teaching ability, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.
4. For intervention tasks, instability is defined as fraction of fails after the first success at each difficulty level by each skill type.
5. Standard errors are reported in parentheses and clustered at village level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.16: The Effects of Interactions on the Instability of Gross Motor Tasks at Each Level (IV)

	Gross Motor Task Difficulty Levels				
	5	6	7	8	9
Interaction Quality:	-0.220*	-0.374*	-0.086	-0.186**	0.002
Home Visitor and Caregiver	(0.131)	(0.192)	(0.066)	(0.082)	(0.063)
Interaction Quality:	0.015	0.022	-0.012	-0.015	-0.011
Home Visitor and Child	(0.061)	(0.038)	(0.012)	(0.027)	(0.024)
Teaching Ability	0.117	0.460***	-0.032	0.279**	-0.020
	(0.163)	(0.170)	(0.080)	(0.112)	(0.139)
Grandmother Rearing[1]	-0.029	-0.172	0.039	0.036	-0.006
	(0.162)	(0.151)	(0.048)	(0.092)	(0.055)
Monthly Age	0.068	-0.003	-0.002	-0.003	-0.006
	(0.063)	(0.014)	(0.002)	(0.008)	(0.004)
Constant	-0.717	0.304	0.152***	0.253	0.267*
	(0.981)	(0.259)	(0.050)	(0.203)	(0.143)
Cragg-Donald F	13.109	13.245	32.846	33.111	28.413
Kleibergen-Paap LM	44.281	19.333	59.405	81.066	43.100
Hansen J	1.650	0.336	2.387	1.120	2.591
P-Value(Hansen J)	0.438	0.846	0.303	0.571	0.274

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, interaction quality between home visitor and caregiver (child) are latent factors based on the supervisor recorded measures. See Appendix C.

4. The instrumental variables include mean, max, and min of other village interaction measures through the same visitor.

5. For intervention tasks, instability is defined as fraction of fails after the first success at each difficulty level by each skill type.

6. For the first stage, we report Crag-Donald F statistics and Kleinbergen-Paap LM statistics. For overidentification test, we report Hansen J statistic and the p-value of Hansen J statistic.

7. Standard errors are reported in parentheses and clustered at village level.

8.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.17: The Effects of Interactions on the Frequency of the Caregiver Playing with the Child by Ability Group

	Fast		Normal		Slow	
	OLS	IV	OLS	IV	OLS	IV
Interaction Quality: Home Visitor and Caregiver	0.058	0.117	0.015	0.023	-0.016	0.081^{**}
	(0.035)	(0.080)	(0.013)	(0.017)	(0.022)	(0.033)
Interaction Quality: Home Visitor and Child	-0.001	0.003	0.002	-0.001	-0.001	-0.007
	(0.007)	(0.015)	(0.002)	(0.003)	(0.004)	(0.007)
Teaching Ability	0.024	0.026	-0.004	-0.009	0.063^{*}	0.051
	(0.035)	(0.062)	(0.010)	(0.014)	(0.034)	(0.066)
Monthly Age	0.003^{**}	0.003	0.001^{*}	0.000	0.002	0.003^{**}
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Grandmother Rearing	0.030	0.021	0.010	0.017^{**}	0.020	0.044
	(0.023)	(0.028)	(0.006)	(0.007)	(0.028)	(0.056)
Constant	5.261^{***}	5.258^{***}	5.465^{***}	5.515^{***}	5.597^{***}	5.597^{***}
	(0.043)	(0.082)	(0.018)	(0.018)	(0.044)	(0.033)
Sargan-Hansen statistic		3.203		1.395		6.016
P-value of Sargan-Hansen statistic		0.361		0.707		0.111
First Stage F value		620.129		2176.970		459.326

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 instrumental variables include mean, max, and min of other village interaction measures using the same visitor.

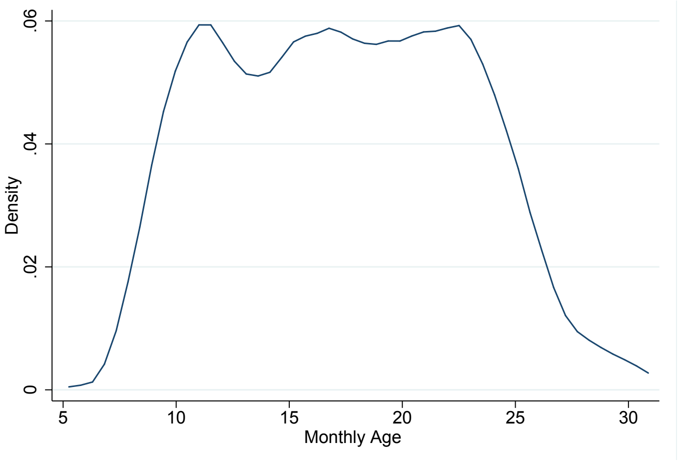
3. Fast group: the children who pass the first task at over 80% of the difficulty levels with an average passing rate greater than 80%. Normal group: the children who do not pass the first task with an average passing rate greater than 80%; or the children pass the first task, and the passing rate is between 50% and 80%. Slow group: the average passing rate is less than 50%.

4. Standard errors in parentheses are clustered at village level.

5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

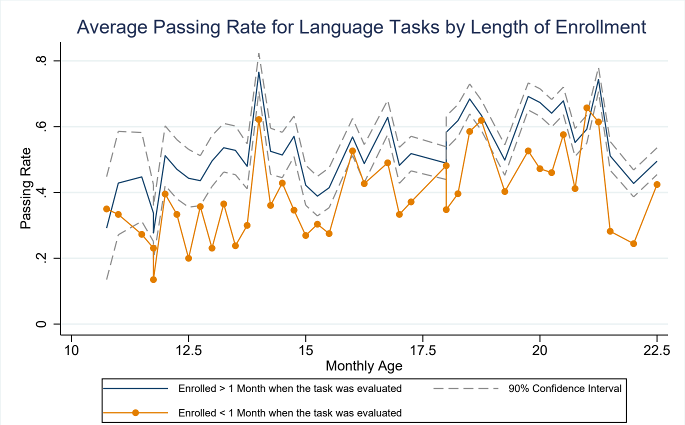
E Nonparametric Tests of a Version of Dynamic Complementarity: Additional Tests

Figure E.1: The Distribution of Monthly Ages When Enrolled into the Program



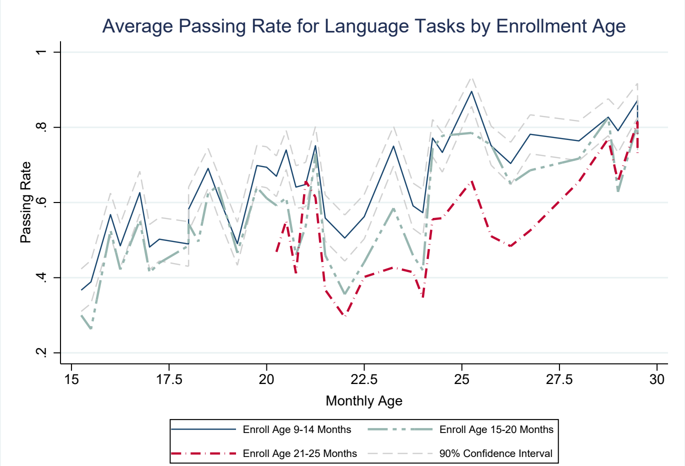
E.1 Language

Figure E.2: Language Tasks Performance Comparison by Length of Enrollment



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years.
3. 90% confidence intervals are shown for both groups.
4. Tasks with fewer than 10 observations in either group are omitted.

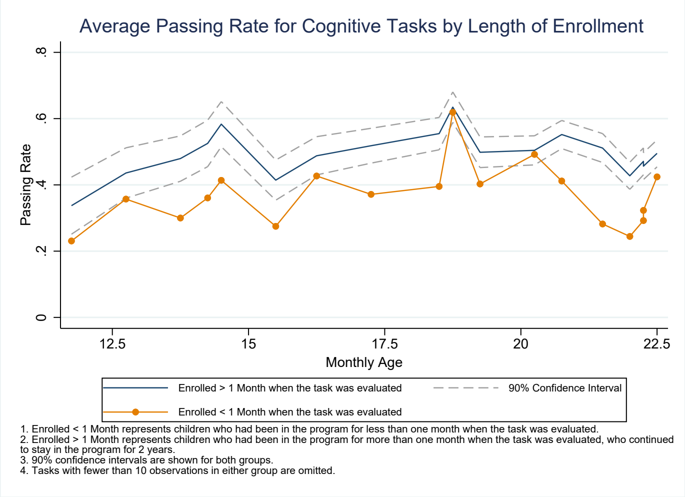
Figure E.3: Average Passing Rate for Language Tasks by Enrollment Age



1. 90% confidence intervals are shown for the groups whose enrollment age is 9-14 months

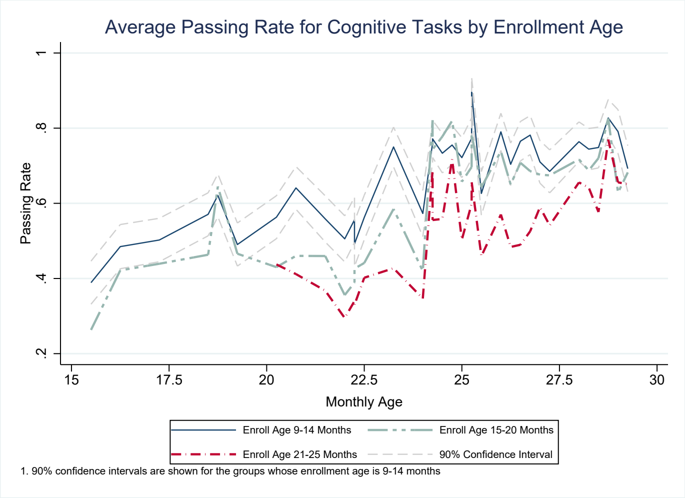
E.2 Cognitive

Figure E.4: Cognitive Tasks Performance Comparison by Length of Enrollment



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years.
3. 90% confidence intervals are shown for both groups.
4. Tasks with fewer than 10 observations in either group are omitted.

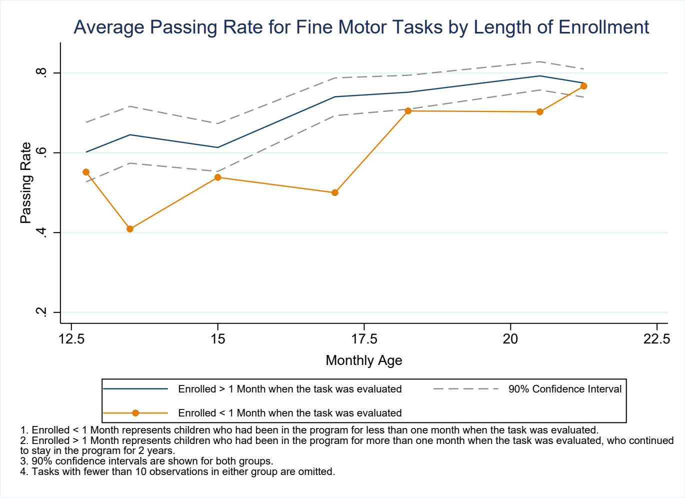
Figure E.5: Average Passing Rate for Cognitive Tasks by Enrollment Age



1. 90% confidence intervals are shown for the groups whose enrollment age is 9-14 months

E.3 Fine Motor

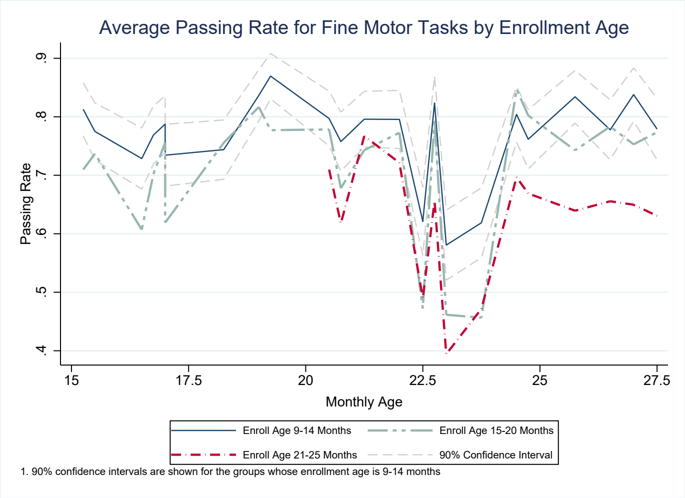
Figure E.6: Fine Motor Tasks Performance Comparison by Length of Enrollment



Average Passing Rate for Fine Motor Tasks by Length of Enrollment

1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years.
3. 90% confidence intervals are shown for both groups.
4. Tasks with fewer than 10 observations in either group are omitted.

Figure E.7: Average Passing Rate for Fine Motor Tasks by Enrollment Age

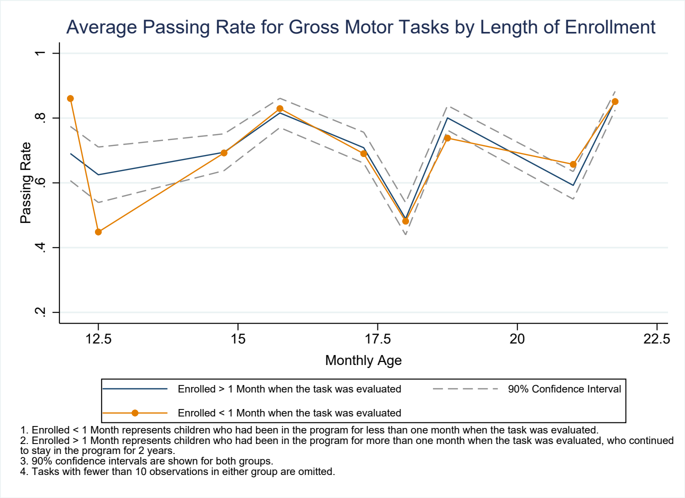


Average Passing Rate for Fine Motor Tasks by Enrollment Age

1. 90% confidence intervals are shown for the groups whose enrollment age is 9-14 months

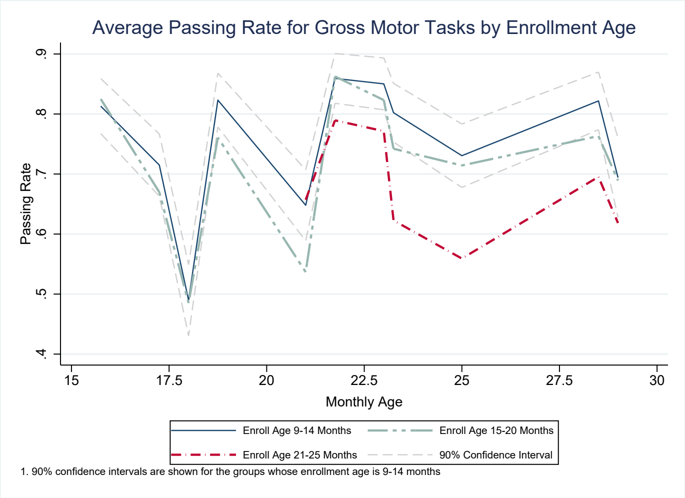
E.4 Gross Motor

Figure E.8: Gross Motor Tasks Performance Comparison by Length of Enrollment



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years.
3. 90% confidence intervals are shown for both groups.
4. Tasks with fewer than 10 observations in either group are omitted.

Figure E.9: Average Passing Rate for Gross Motor Tasks by Enrollment Age



1. 90% confidence intervals are shown for the groups whose enrollment age is 9-14 months

F Identification

To simplify the notation, we suppress skill index s and analyze a model for a single skill. The same framework applies to all skills. Skill at level ℓ can be written as:

$$\begin{aligned}\ln K(\ell, a) =& \ln K(\ell, a-1) + \delta_\ell(a)\eta + V_\ell(Q(a)) + \varepsilon_\ell(a), \quad \underline{a}(\ell) \leq a \leq \bar{a}(\ell) \\ E(\varepsilon_\ell(a)) =& 0, \quad \text{Var}(\varepsilon_\ell(a)) = \sigma^2_{\varepsilon(\ell)}. \end{aligned} \tag{F.1}$$

At the transition point from $\ell - 1$ to ℓ, under skill invariance the equation can be written, at the start of ℓ, as:

$$\begin{aligned}\ln K(\ell, \underline{a}(\ell)) =& \ln K(\ell-1, \bar{a}(\ell-1)) + \delta_\ell(\underline{a}(\ell))\eta + V_\ell(Q(\underline{a}(\ell))) + \varepsilon_\ell(a(\ell)), \\ & \underline{a}(\ell) \leq a \leq \bar{a}(\ell). \end{aligned} \tag{F.2}$$

Define $\Delta_\ell(a) := \sum_{j=\underline{a}(\ell)}^{a} \delta_\ell(j); \quad \Lambda_\ell(a) := \sum_{k=\underline{a}(\ell)}^{a} V_\ell(Q(k))$, $U_\ell(a) := \sum_{j=\underline{a}(\ell)}^{a} \varepsilon_\ell(j)$.

We assume $\delta_\ell(a) = \delta_\ell$ (for all a in ℓ) is consistent with the same skill being taught at each level. We assume η (ability) is constant across all levels, and we parameterize the model such that

$$\eta\delta_\ell(a) = \bar{\beta}_\ell(X) + \omega_\ell$$

where X includes determinants of ability and learning such as family background and interaction measures. In this notation, $\eta\Delta_\ell(a) := (a - \underline{a}(\ell))(\bar{\beta}_\ell(X) + \omega_\ell)$. We assume that $\omega_\ell = \omega$ for all ℓ, $E(\omega) = 0$, $\text{Var}(\omega) = \sigma^2_\omega$. We assume normal errors in making our estimates. Thus, we restrict attention to identification of means and covariances.

F.1 Recursive Definition of the Skill Index

When $\underline{a}(1) = 0$, level 1 skill is as following:

$$\ln K(1,a) = \mu_1 + \mu_0(Z) + V(Q(a)) + \bar{\beta}_1(X)a + \underbrace{\left\{ a\omega + \sum_{j=1}^{a} \varepsilon_1(j) + \Upsilon \right\}}_{\Psi_1(a)}. \tag{F.3}$$

It is useful to collect the unobservables into

$$\Psi_1(a) = \{a\omega + U_1(a) + \Upsilon\}. \tag{F.4}$$

We assume (ω, Υ) are mutually independent, that $\varepsilon_\ell(a)$ is independent of $\varepsilon_{\ell'}(a')$, $\ell' \neq \ell$, and $(\omega, \Upsilon) \perp\!\!\!\perp \varepsilon_\ell(a)$ for all ℓ, a.

$$\mathrm{Var}(\Psi_1(a)) = a^2\sigma_\omega^2 + a\sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2 \tag{F.5}$$

and we define $\mathrm{Var}(\Psi(a)) := \sigma^2(1,a)$.

$$\mathrm{Cov}(\Psi,(a),\Psi,(a')) = aa'\sigma_\omega^2 + \min(a,a')\sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2. \tag{F.6}$$

F.2 Identification of First Level Parameters

Under conditions presented in Matzkin (1993, 2007), we can identify up to scale $\sigma(1,1)$. This follows from standard results in the binary choice model, provided that at least one parameter is constant over the ℓ interval.

In estimation, we assume $V_\ell(\Lambda_\ell(a)) := \sum_{j=0}^{J} \lambda_j a^j$ to capture aging and maturation effects, which we assume operates uniformly across ℓ. We assume $\lambda_j = 0, j > 2$ (linear model). Although in principle we could identify ℓ-specific maturation effects, we do not do it here. Shocks are i.i.d. with mean zero (i.e., $E(\varepsilon_\ell(a)) = 0$) and with variance $\sigma^2_{\varepsilon(\ell)}$.

In each interval of ℓ, $\underline{a} \leq \ell \leq \bar{a}(\ell)$.

$$\ln K(\ell, a) = \eta \Delta_\ell(a) + \Lambda_\ell(a) + U_\ell(a) + \ln K(\ell, \underline{a}(\ell)). \tag{F.7}$$

Define $\bar{K}_\ell$ as the minimum level of mastery of skill ℓ.

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

$$\Pr(D(\ell,a) = 1) = \Pr(\ln K(\ell,a) \geq \bar{K}_\ell) = \Pr(\ln K(\ell, \underline{a}(\ell)) + \eta\Delta_\ell(a) + \Lambda_\ell(a) + U_\ell(a) \geq \bar{K}_\ell).$$

The introduction of $\bar{K}_\ell$ in the estimation adds an intercept to the model that is the same for each a in ℓ, but in general differs across s.

For $\ell = 1$, we characterize the initial condition by $K(1, \underline{a}(1)) = \mu_0(Z) + \Upsilon$,

$E(\Upsilon) = 0$, $\text{Var}(\Upsilon) = \sigma^2_{\Upsilon}$, for the interval $[\underline{a}(1), \bar{a}(1)]$.

Thus, at $\underline{a}(1) = 1$, we can identify

$$\frac{\mu_1}{\sigma(1,1)}, \frac{\mu_0(Z)}{\sigma(1,1)}, \frac{\lambda_0}{\sigma(1,1)}, \frac{\lambda_1}{\sigma(1,1)}, \frac{\bar{\beta}_1(X)}{\sigma(1,1)}.$$

μ_1 is the intercept which includes $\bar{K}_1$. At the next age, we can identify

$$\frac{\mu_1}{\sigma(1,2)}, \frac{\mu_0(Z)}{\sigma(1,2)}, \frac{\lambda_0}{\sigma(1,2)}, \frac{\lambda_1}{\sigma(1,2)}, \frac{\bar{\beta}(X)}{\sigma(1,2)}.$$

Invoking the constancy of at least one parameter across age 1 and 2, we can identify

$$\frac{\sigma(1,2)}{\sigma(1,1)}$$

and using the same logic across ages for $\ell = 1$, we can identify

$$\frac{\sigma(1,j)}{\sigma(1,1)}, \qquad \underline{a}(1) \leq j \leq \bar{a}(1).$$

We can clearly identify $\bar{K}_\ell$, $\ell = 1, \ldots, L$ up to scale and an unknown constant if we assume intercepts are constant across levels.

F.3 Identification of Variance and Covariance Terms at the First Level

From (F.7), we can identify $\sigma^2_\omega, \sigma^2_{\varepsilon(1)}, \sigma^2_\Upsilon$ up to $\sigma^2(1,1)$ for $\underline{a}(1) \leq a \leq \bar{a}(1)$. Following Carneiro et al. (2003) and Heckman and Vytlacil (2007b) from the joint probabilities

$$\begin{aligned}&\Pr(D(1,a) = d(a)), \quad \Pr(D(1,a') = d(a'))\\ &d(a), d(a') \in \{0,1\}\end{aligned}$$

we can identify

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

See Heckman and Vytlacil (2007a).

F.4 Identification of Higher Level Parameters

The same logic extends to higher levels of ℓ, $L \geq \ell > 1$, except here it is fruitful to distinguish two cases: with and without skill invariance. We first assume skill invariance over levels. We maintain skill invariance *within* the same level of skill.

Under Skill Invariance Across Levels

We can write $K(\ell, a)$ as

$$\begin{aligned}\ln K(\ell, a) =& \mu_0(Z) + \sum_{k=1}^{\ell-1} \bar{\beta}_k(X)(\bar{a}(k) - \underline{a}(k)) + \bar{\beta}_\ell(X)(a - \underline{a}(\ell)) + \sum_{k=1}^{\ell-1} \Delta_k(\bar{a}(k)) + \Delta_\ell(a) \\ &+ \sum_{k=1}^{\ell-1} \Lambda_k(Q(\bar{a}(k))) + \Lambda_\ell(a) + \underbrace{\left\{ (a - \underline{a}(\ell))\omega + \sum_{k=1}^{\ell-1} U_k(\bar{a}(k)) + U_\ell(a) + \Upsilon \right\}}_{\Psi_\ell(a)}.\end{aligned}$$

For each a in each level ℓ, we acquire the threshold $\bar{K}_\ell$ as an intercept term in the sequence of observed indicators $D(\ell, a)$.

$$\mathrm{Var}\Psi_\ell(a) := \sigma^2(\ell, a) = (a - \underline{a}(\ell))^2 \sigma_\omega^2 + \sum_{k=1}^{\ell-1} \sigma_{\varepsilon(k)}^2 (\bar{a}(k) - \underline{a}(k)) + (a - \underline{a}(\ell)) \sigma_{\varepsilon(\ell)}^2 + \sigma_\Upsilon^2$$

with covariance

$$\begin{aligned}\mathrm{Cov}(\Psi_\ell(a), \Psi_\ell(a')) = & (a - \underline{a}(\ell))(a' - \underline{a}(\ell))\sigma_\omega^2 + \sum_{k=1}^{\ell-1} \sigma_{\varepsilon(k)}^2 (\bar{a}(k) - \underline{a}(k)) \\ &+ \min(a - \underline{a}(\ell), a' - \underline{a}(\ell)) \sigma_{\varepsilon(\ell)}^2 + \sigma_\Upsilon^2.\end{aligned} \tag{F.8}$$

From each indicator variable, we can identify for each ℓ and a, the threshold variables $\frac{\bar{K}_\ell}{\sigma(\ell,a)}$ and $\frac{\mu_0(Z)}{\sigma(\ell,a)}$, $\frac{\bar{\beta}_\ell(X)}{\sigma(\ell,a)}$, $\frac{\Delta_\ell(a)}{\sigma(\ell,a)}$, $\frac{\Lambda_\ell(a)}{\sigma(\ell,a)}$. The other terms in the index $K(\ell, a)$ are identified by a recursive argument starting from $\ell = 1$ (previously discussed). In Equation (F.8), the only unknown parameter is $\sigma_{\varepsilon(\ell)}^2$, and so from it, we can identify the variance term $\sigma_{\varepsilon(\ell)}^2$. Therefore, we can identify the variance of the sum of shocks (i.e., $\mathrm{Var}\Psi_\ell(a)$) at level ℓ at age a.

From the above discussion, we can identify at all levels, $\ell \geq 2$, scales of variance $\text{Var}\Psi_\ell(a)$ without imposing additional normalizations. The only normalization we need is the scale of variance term $\sigma(1, j) = 1$ at level one.

Without Skill Invariance

To test the skill invariance assumption, we use an affine transformation as follows:

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

It is clearly impossible to separate $\gamma_{0,\ell}$ from the threshold parameter $\ln \bar{K}(\ell)$ so we normalize $\gamma_{0,\ell} = 0$ with the understanding that any estimated threshold parameter is net of $\gamma_{0,\ell}$ (i.e., $\bar{K}_\ell - \gamma_{0,\ell}$). To identify the parameter $\gamma_{1,\ell}$, we use additional moments. Using the definition of $\bar{\Psi}_\ell(a)$, we obtain:

$$\begin{aligned}\text{Cov}\big(\Psi_\ell(a), \Psi_{\ell-1}(a')\big) = \gamma_{1,\ell} \Bigg\{ &(a - \underline{a}(\ell))(a' - \underline{a}(\ell - 1))\sigma_\omega^2 + \sigma_\Upsilon^2 + \sum_{k=1}^{\ell'-2} (\bar{a}(k) - \underline{a}(k))\sigma_{\varepsilon(k)}^2 \\ &+ (a' - \underline{a}(\ell' - 1))\sigma_{\varepsilon(\ell-1)}^2 \Bigg\}. \end{aligned} \tag{F.9}$$

we can identify $\gamma_{1,\ell}$ based on this covariance term between level ℓ and $\ell - 1$. Using similar logic, we can identify the parameters $\gamma_{1,\ell}$ across difficulty levels. Thus, this aspect of skill invariance can be tested.

Beyond Normality

Drawing on the analysis of Heckman and Vytlacil (2007a) and Matzkin (1993, 2007), with sufficient variation in the regressors, we can not only identify the model under normality, but we can also identify the model under more general distributional assumptions based on Matzkin (1993).

To identify all the shock distributions, we first need to nonparametrically identify the distribution of the sum of all shocks (i.e.,$\bar{\Psi}_1(a)$). We need to impose assumptions on $\mu_1 + \mu_0(Z) + V(Q(a)) + \bar{\beta}_1(X)a$ and the shock terms. We need (a) the function is continuous; (b) there exists some x' and z' such that $\mu_1 + \mu_0(z') + V(Q(a)) + \bar{\beta}_1(x')a = R$ regardless of the specified coefficients (c) the shock term $\bar{\Psi}_1(a)$ and the observable covariates are independent. (d) the distributions of all shocks are continuous. These assumptions guarantee nonparametric identification of $\bar{\Psi}_1(a)$. Next, we discuss how to separately identify the distributions of ω, ε and Υ. We use characteristic functions.

Denote the characteristic function of $\bar{\Psi}_1(a)$ by $\varphi_{\Psi(a)}(t)$. Since we assume that all the shock terms are independent with each other, we have the following condition: $\varphi_{\Psi(a)}(t) = \varphi_\omega(at)\varphi_{U_1(a)}(t)\varphi_\Upsilon(t)$ Notice that $U_1(a) = \varepsilon_1(1) + \cdots + \varepsilon_1(a)$. Therefore, $\varphi_{U_1(a)}(t) = \varphi_{\varepsilon(1)}(t)^a$. Thus, we obtain the following equation:

$$\varphi_{\Psi(a)}(t) = \varphi_\omega(at)\varphi_{\varepsilon(1)}(t)^a\varphi_\Upsilon(t).$$

If we assume that ω is standard normal, we can identify

$$\varphi_\Upsilon(t) = \frac{\varphi_{\Psi(2)}(t)/\varphi_\omega(2t)}{\varphi_{\Psi(1)}(t)/\varphi_\omega(t)}$$

Collecting results, the characteristic function for ε_1 can be identified from

$$\varphi_{\varepsilon_{(1)}}(t) = \frac{\varphi_{\Psi(1)}(t)}{\varphi_{\omega}(t)\varphi_{\Upsilon}(t)}.$$

Since we can nonparametrically identify the shock distributions, we can easily get the moments of the first two levels. Identification of the remaining parameters follows from repeated application of the same logic.

G Simulation Procedure for Method of Moments Estimation

Our simulation procedure is as follows. We parameterize each function and make the distributional assumptions noted below. We adopt a new subscript notation unique to this appendix to simplify the notation and distinguish person-specific variables. Thus, we write $\varepsilon_\ell(s,a)$ for person i as $\varepsilon_i(s,\ell,a)$ with similar notational changes for i-subscripted variables.

(A) We first simulate the initial conditions for latent skills: $\ln K(s,0) = Z_i'\beta_{0,s} + \tau_{i,0,s}$ for each child i, where Z_i is a vector of variables for child i including family background measures and the monthly age of the child at enrollment. Family background information includes father's years of education, mother's years of education, and grandmother's years of education. $\tau_{i,0,s}$ is a random term independent of all regressors and error terms at all levels. It follows the normal distribution $N(0,\sigma_\tau^2(0,s))$.

(B) After drawing initial latent skills $K(s,0)$, child i skills evolve following the equation $(\ln K_i(s,\ell,a) = \ln K_i(s,\ell,a-1) + (\delta_\ell(s)\eta_i(s) + V(Q(a)) + \varepsilon_i(s,\ell,a)))$. In this equation, $\delta_\ell(s)$ is a level-specific lesson parameter for skill s. $\varepsilon_i(s,\ell,a)$ is an i.i.d. random term associated with the learning process drawn from a normal distribution $N(0,\sigma^2(s,\ell))$ for skill s, and $\eta_i(s)$ is an individual learning parameter with the specification: $\eta_i(s) = X\beta_s + \gamma_i$, where X includes the set of interaction measures; and γ_i is individual specific learning ability, which follows the normal distribution $N(0,\sigma_{s,\gamma}^2)$.

Since $\delta_\ell(s)$ and $\eta_i(x(a))$ enter the model multiplicatively, we cannot separately identify them without some normalization. We normalize the first level $\delta(s,1) = 1$ to identify the coefficients of β_s. Since we assume that the β_s are the same across levels, we can identify $\delta(s,\ell)$ for $\ell > 1$. $\delta(s,\ell)\eta_i(x(a)) = \delta(s,\ell)(X'(a)\beta_s + \gamma_i)$ where the X includes determinants of ability and $\delta(s,\ell)$ captures lesson contents. We now give an example on how to simulate the latent skills within the same difficulty level ℓ:

1. Use the initial latent skill $\ln K_i(s,0)$ which is formed by $Z_i'\beta_{0,s} + \tau_{i,0,s}$ and the random draw $\tau_{i,0,s}$ from a normal distribution with mean zero and variance $\sigma_\tau^2(0,s)$. Child ages sometimes differ at enrollment.
2. If child i starts the intervention at difficulty level ℓ, we randomly draw the task error term at difficulty level ℓ from the normal distribution with mean zero and variance $\sigma^2_{\varepsilon(\ell)}$. Then, we can construct the latent skill for the first task based on the following equation:
   $\ln K_i(s,\ell,1) = \ln K(s,0) + \delta(s,\ell)(X'(1)\beta_s + \gamma_i) + \varepsilon_i(s,\ell,1)$
3. Similarly we randomly draw a shock $\varepsilon_i(s,\ell,2)$ for the 2nd task at level ℓ, and then, we construct the latent skill following:
   $\ln K_i(s,\ell,2) = \ln K_i(s,\ell,1) + \delta(s,\ell)(X'(2)\beta_s + \gamma_i) + \varepsilon_i(s,\ell,2)$
4. Repeat step 3 until the last task of difficulty level ℓ at age $\bar{a}(s,\ell)$ to construct the latent skill:
   $\ln K_i(s,\ell,j) = \ln K_i(s,\ell,j-1) + \delta(s,\ell)(X'(j)\beta_s + \gamma_i) + \varepsilon_i(s,\ell,j)$

(C) Under skill invariance, the latent skill across difficulty levels evolves as follows:

$$\begin{aligned}\ln K_i(s,\ell+1,\underline{a}(s,\ell+1)) &= \ln K_i(s,\ell,\bar{a}(s,\ell)) + \delta(s,\ell+1)(X'(a)\beta_s+\gamma_i)\\ &\quad + \varepsilon_i(s,\ell+1,a(s,\ell+1)).\end{aligned}$$

After the first task at that level, latent skills evolve as follows for $j>1$:

$$\ln K_i(s,\ell+1,a) = \ln K_i(s,\ell+1,a-1)+\delta(s,\ell+1)(X'(a)\beta_s+\gamma_i)+\varepsilon_i(s,\ell+1,j).$$

(D) For a model without skill invariance, the latent skill across difficulty levels develops as follows:
$\ln K_i(s,\ell+1,a(s,\ell,1)) = \gamma_{0,\ell} + \gamma_{1,\ell}(\ln K_i(s,\ell)) + \delta(s,\ell+1)(X'(1)\beta_s+\gamma_i) + \varepsilon_i(s,\ell+1,a(s,\ell+1))$.
After the first task, latent skill on other tasks at level $\ell+1$ are as follows:
$\ln K_i(s,\ell+1,a(s,\ell+1,j)) = \ln K_i(s,\ell+1,a(s,\ell+1,j-1))+\delta(s,\ell+1)(X'(j)\beta_s+\gamma_i)+\varepsilon_i(s,\ell+1,j))$, $j>1$.

(E) Given estimates of the parameters $\bar{K}(s,\ell)$, we can calculate the simulated task performance based on the following equation:

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

(F) We form moments based on series of simulated child task performance and minimize the distance between the simulated and the data moments.

(G) The moments we consider in estimation include: (1)All task passing rates; (2) The passing rate on the first five tasks at each level; (3) The passing rate for each difficulty level; (4) The passing rate for newly enrolled children; (5) The passing rate for those who enroll in the program longer than one month; (6) The probability of passing the j'th task ($j \neq j'$) within each level, conditioning on the child passing the jth task; (7) The probability of passing the j'th task at level $\ell + 1$ across all difficulty levels, conditioning on the child passing the jth task at level ℓ; (8) The probability of passing the j'th task at level $\ell + 2$, conditioning on the child passing the jth task at level ℓ.

(G) After obtaining the point estimates, we bootstrap to calculate the standard errors of the estimates.

Table G.1: Assumptions on Random Shocks in the Model

Parameters	Level Specific or Not	Distribution
Initial Latent Skill Condition Shock ($\tau_{i,0,s}$)	No	Normal$(0, \sigma_{\tau}^2(0, s))$
Learning ability Shock ($\omega_i(s)$)	No	Normal$(0, \sigma_{\omega}^2(0, s))$
Task Performance Shock ($\varepsilon_i(s, \ell, a)$)	Yes	Normal$(0, \sigma_{\varepsilon,s,\ell}^2)$

G.1 Bootstrap Procedure

Since our data are clustered, to conduct robust inference, we use the paired cluster bootstrap method in our paper. In the paper, the clustering is at the village level. We document the paired cluster bootstrap procedure below.

- From the original sample, we get point estimates β^*.

- We iterate the following bootstrap procedure 1000 times from a sample of G clusters $(y_1, \boldsymbol{X}_1), \cdots, (y_G, \boldsymbol{X}_G)$, resampling with replacement G times from the original sample of clusters. The unit of bootstrap is at cluster level. After we randomly draw G clusters, we construct one bootstrap sample.

- Based on the bootstrap sample k, we estimate the structural model based on the estimation procedure documented above and get the point estimates β_k^{bs} for the structural model.

- We conduct inference on each estimated parameter β^* based on the distribution of β^{bs}. After iterating the bootstrap 1000 times, we have the distribution of each parameter. We then calculate the confidence interval and standard error for each parameter.

H Estimation: Moment Fit

Contents

Table H.1: Moments Used in Estimation

	Number of Moments		
Moment	Language	Cognitive	Fine Motor
All task passing rate	103	70	30
The first five task passing rates at each level	50	48	24
Each difficulty level task passing rate	10	13	7
Each task passing rate for newly enrolled children (≤ 1 month)	71	45	14
Each task passing rate for children enrolled in the program for > 1 month	96	70	30
Each difficulty level duration measure	10	13	7
Each difficulty level correlation between duration and interaction measures	30	36	21
Within each level, conditional on children who can pass the jth task, the probability of passing the j'th task ($j \neq j'$)	100	82	43
Across difficulty levels, conditional on children who can pass the jth task at level ℓ, the probability of passing the j'th task at level $\ell + 1$	225	177	84
Across difficulty levels, conditional on children who can pass the jth task at level ℓ, the probability of passing the j'th task at level $\ell + 2$	200	142	79
Total	895	696	339

In summary, 80% of the simulated moments are in the 95% confidence intervals of data moments.

- Overall, our estimates fit the moments very well. The model without skill invariance has better fit.
- χ^2 test results are reported in column χ^2 of Table H.2.
- We also examine the model of fit by the following summary measure:

  $$R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$$

  where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment i. Values of R are reported in column R of Table H.2.

Table H.2: Goodness of Fit Summary

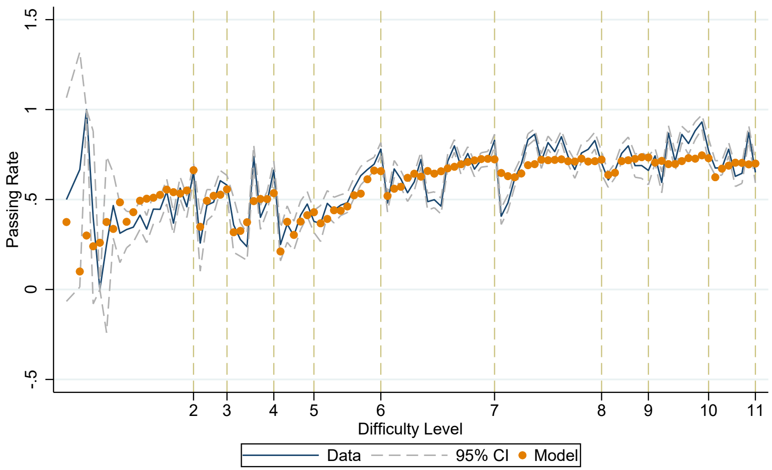
	Language		Cognitive		Fine Motor	
	χ^2	R	χ^2	R	χ^2	R
With skill invariance	32.71	133.05	16.84	93.5	5.59	29.09
Without skill invariance	21.27	121.39	14.31	81.63	5.23	24.73

1. $R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$, where y_i^m is the predicted moment i for the model, and y_i^d is the empirical moment.
2. For all models, we cannot reject the model at the 0.0001 level.

H.1 Language

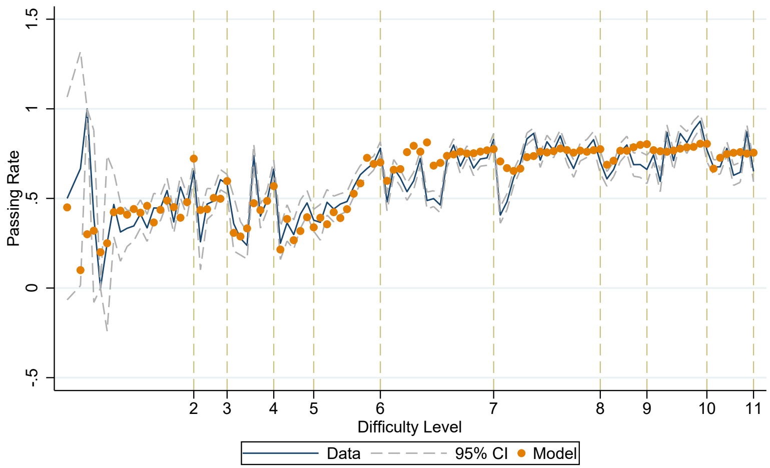
Figure H.1: Fit on All Language Tasks by Level

(a) Skill Invariance



Note: 1. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.
2. All the children started from level 2 or above upon enrolling.

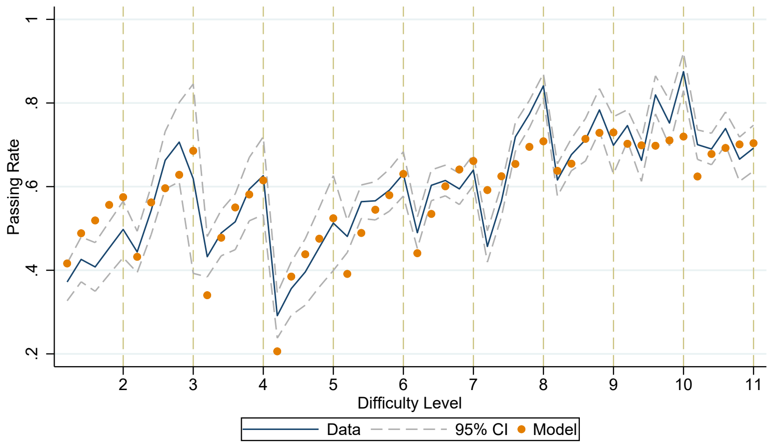
(b) w/o Skill Invariance



Note: 1. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.
2. All the children started from level 2 or above upon enrolling.

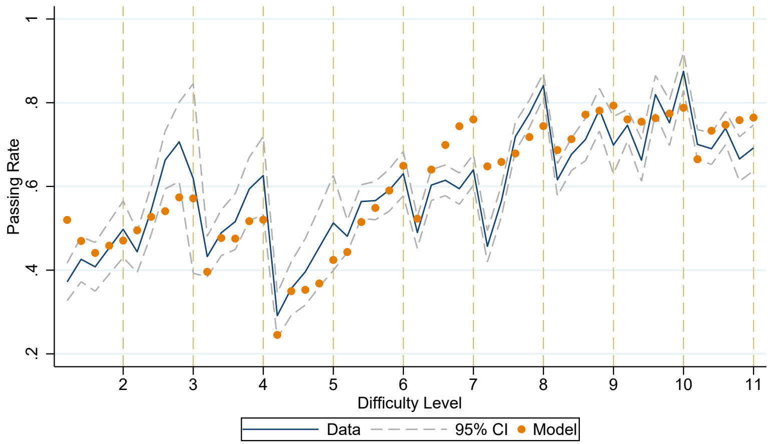
Figure H.2: Fit on the First Five Tasks in Each Level

(a) Skill Invariance



Note: 1. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.
2. All the children started from level 2 or above upon enrolling.

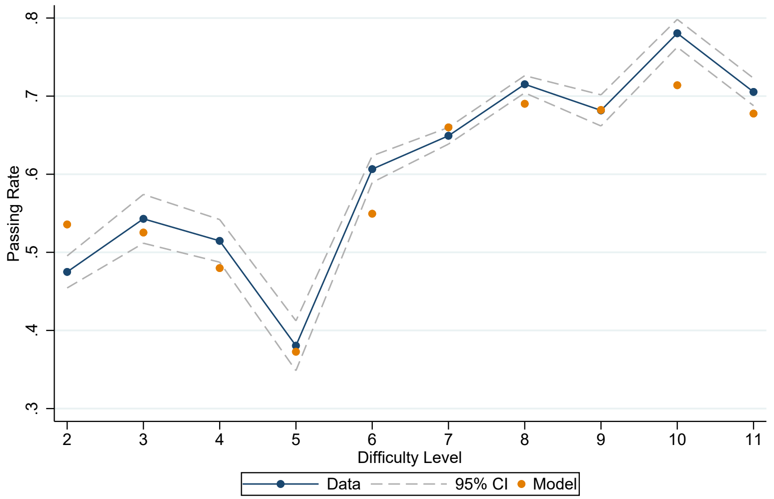
(b) w/o Skill Invariance



Note: 1. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.
2. All the children started from level 2 or above upon enrolling.

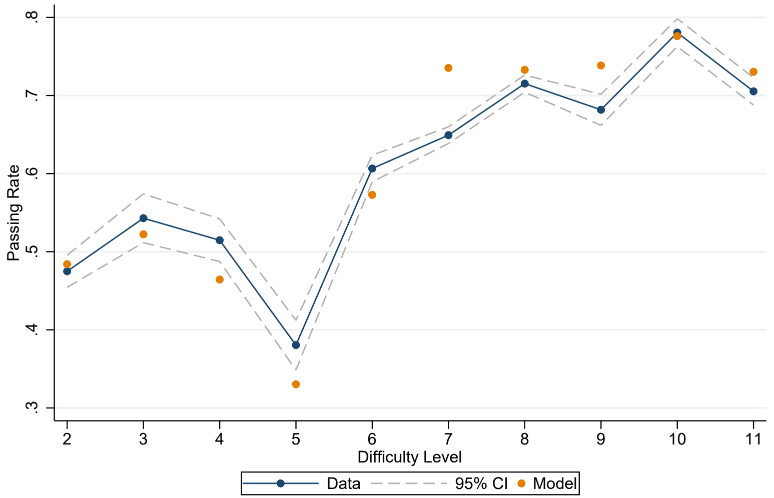
Figure H.3: Fit on Average Passing Rate by Level

(a) Skill Invariance



All the children started from level 2 or above upon enrolling.

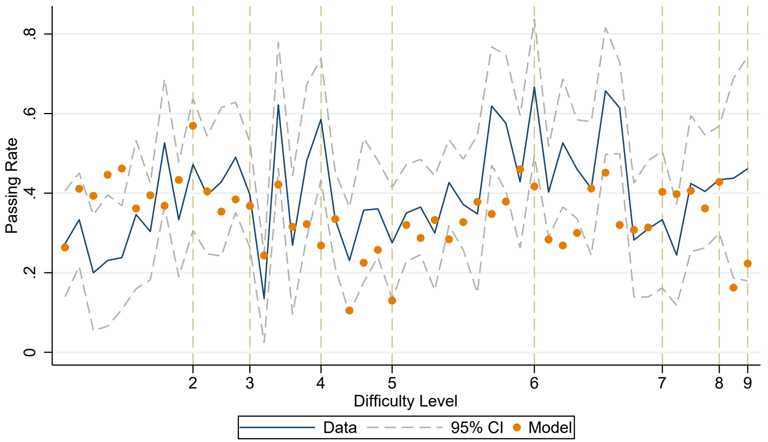
(b) w/o Skill Invariance



All the children started from level 2 or above upon enrolling.

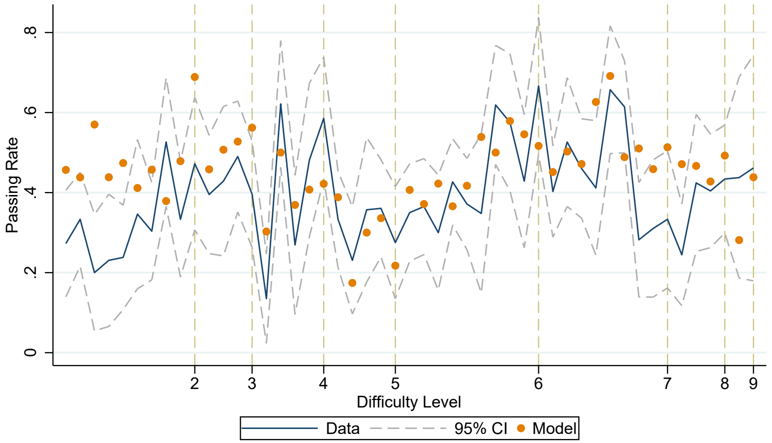
Figure H.4: Fit by Length of Enrollment: Newly Enrolled Group

(a) Skill Invariance



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level . Within difficulty levels, tasks are arranged in the order of the children taking them.
4. All the children started from level 2 or above upon enrolling.

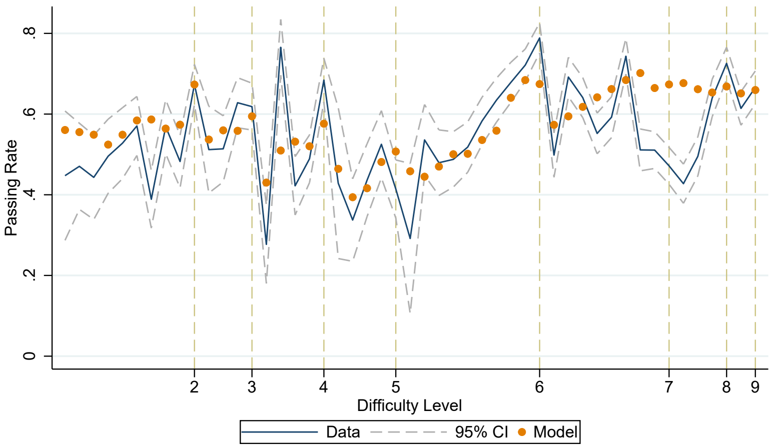
(b) w/o Skill Invariance



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level . Within difficulty levels, tasks are arranged in the order of the children taking them.
4. All the children started from level 2 or above upon enrolling.

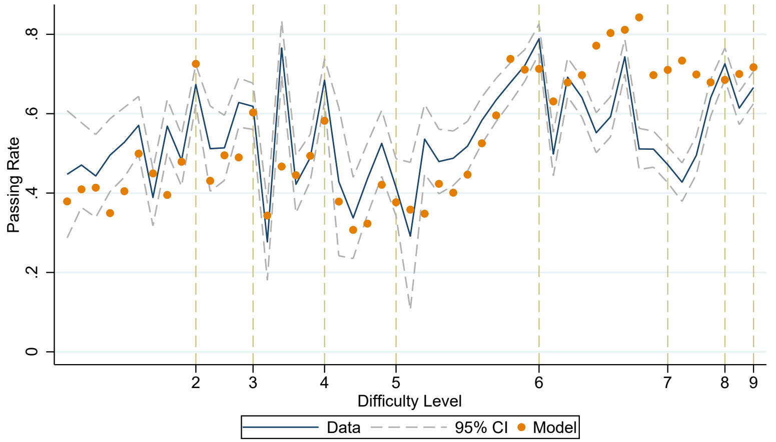
Figure H.5: Fit by Length of Enrollment: Group Enrolled > 1 Month

(a) Skill Invariance



1. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years. 2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them. 4. All the children started from level 2 or above upon enrolling.

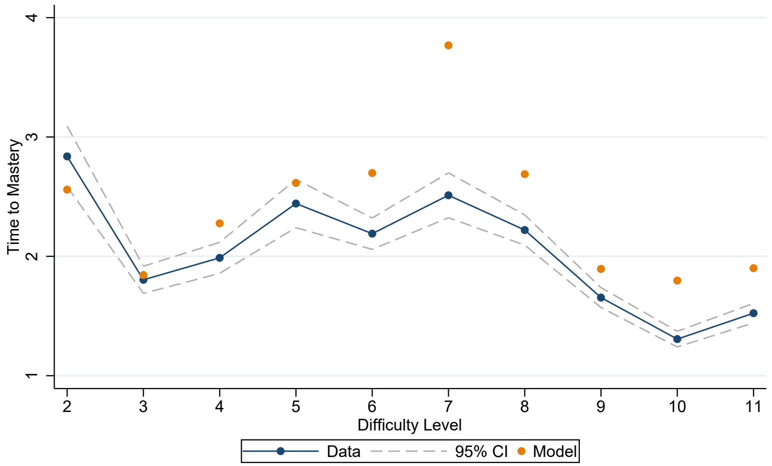
(b) w/o Skill Invariance



1. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years. 2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them. 4. All the children started from level 2 or above upon enrolling.

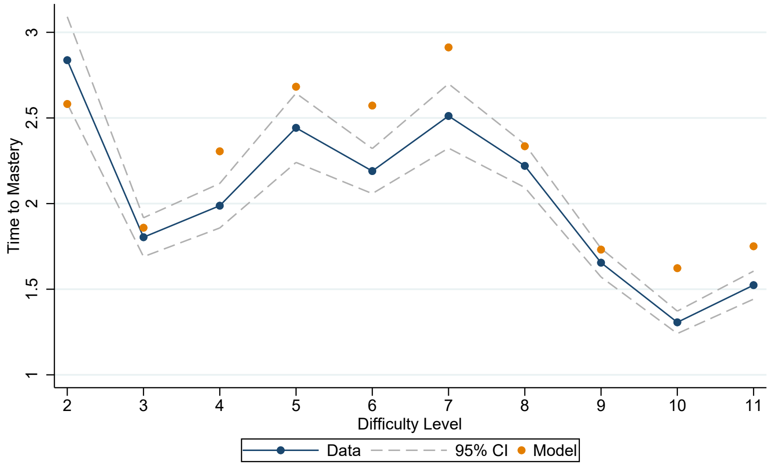
Figure H.6: Fit on Time to Mastery by Level

(a) Skill Invariance



Note: 1. Time to Mastery is defined as the number of tasks a child takes at the previous difficulty level until the first success (inclusive).
2. All the children started from level 2 or above upon enrolling.

(b) w/o Skill Invariance

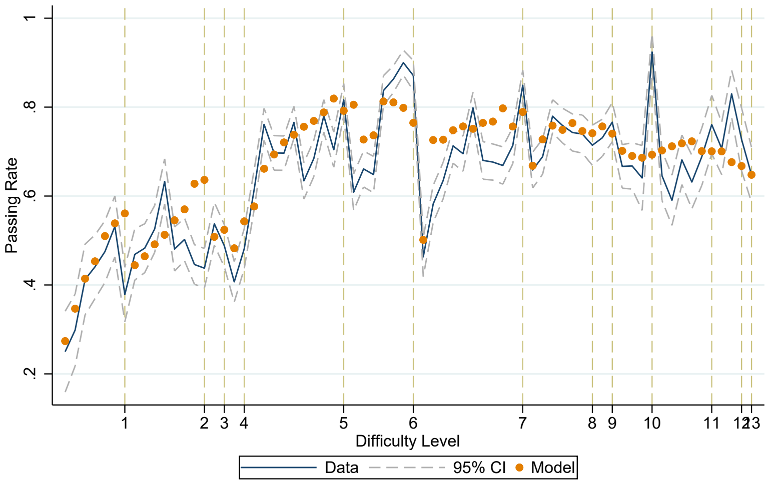


Note: 1. Time to Mastery is defined as the number of tasks a child takes at the previous difficulty level until the first success (inclusive).
2. All the children started from level 2 or above upon enrolling.

H.2 Cognition

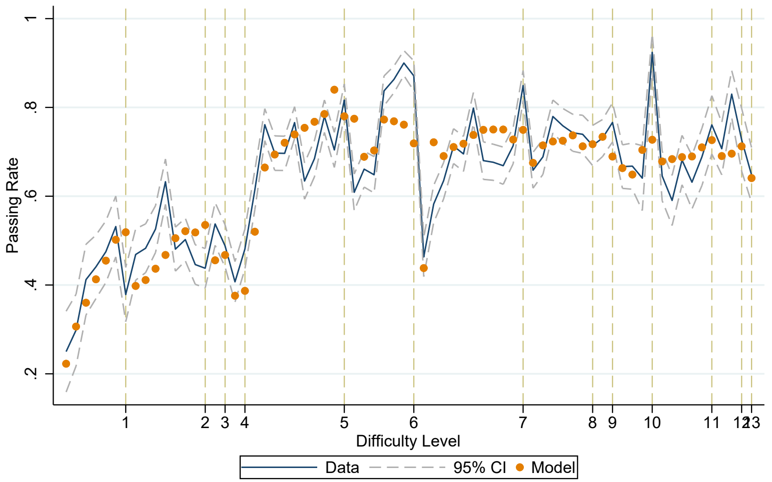
Figure H.7: Fit on All Cognitive Tasks by Level

(a) Skill Invariance



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

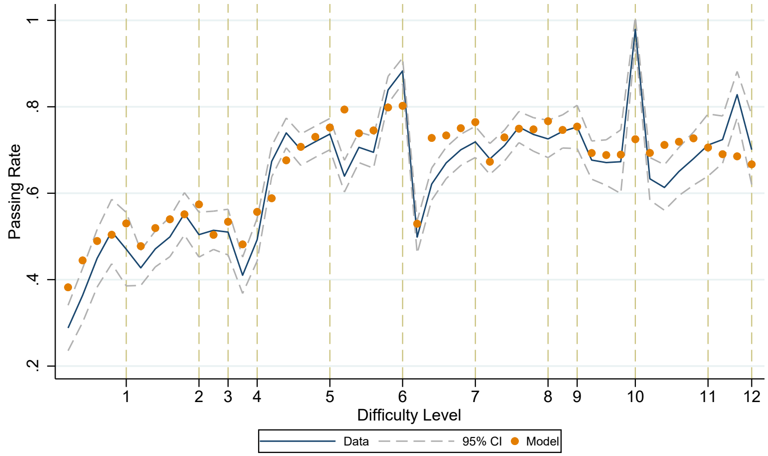
(b) w/o Skill Invariance



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

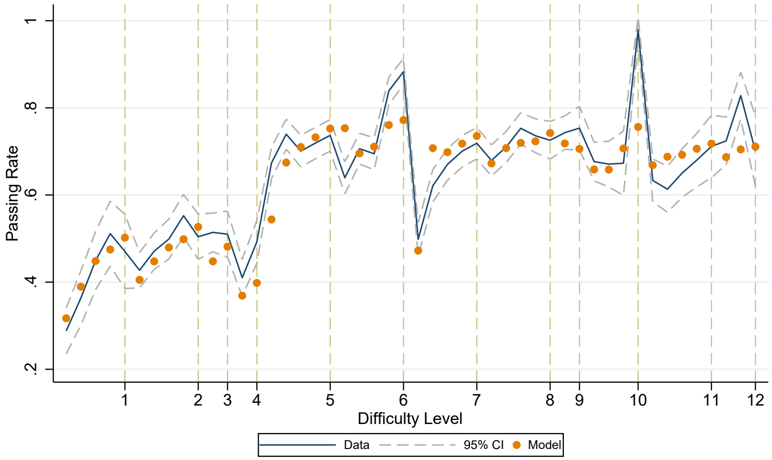
Figure H.8: Fit on the First Five Tasks in Each Level

(a) Skill Invariance



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

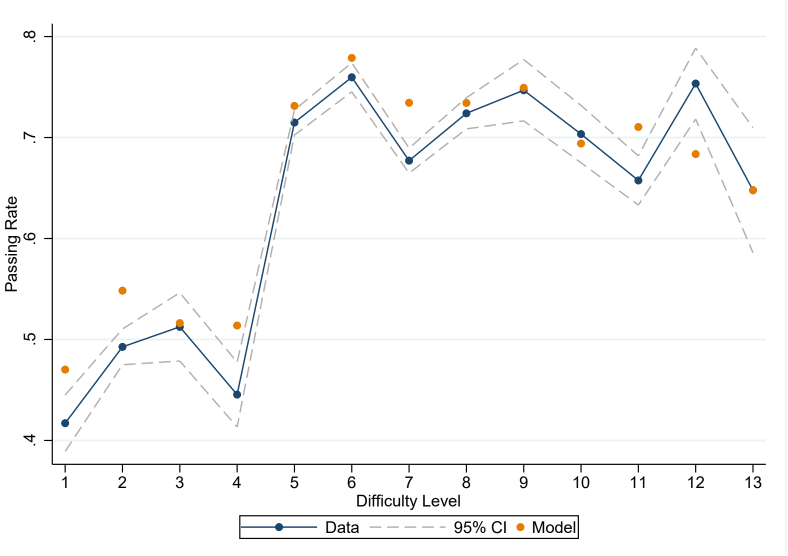
(b) w/o Skill Invariance



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

Figure H.9: Fit on Average Passing Rate by Level

(a) Skill Invariance



(b) w/o Skill Invariance

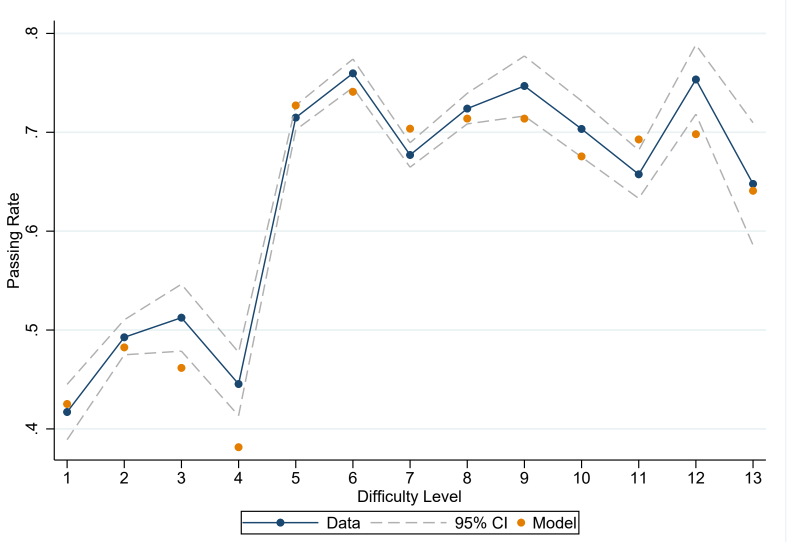
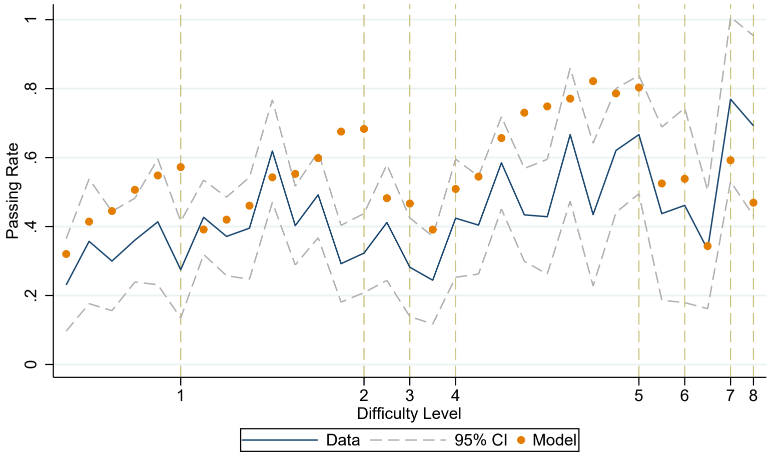


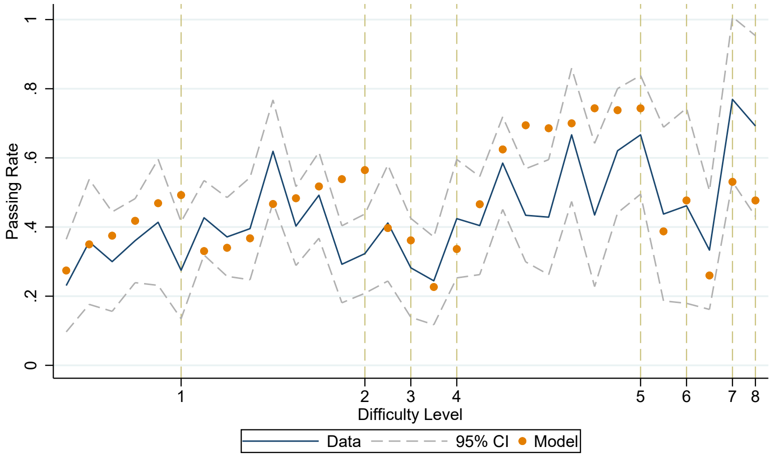
Figure H.10: Fit by Length of Enrollment: Newly Enrolled Group

(a) Skill Invariance



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level . Within difficulty levels, tasks are arranged in the order of the children taking them.

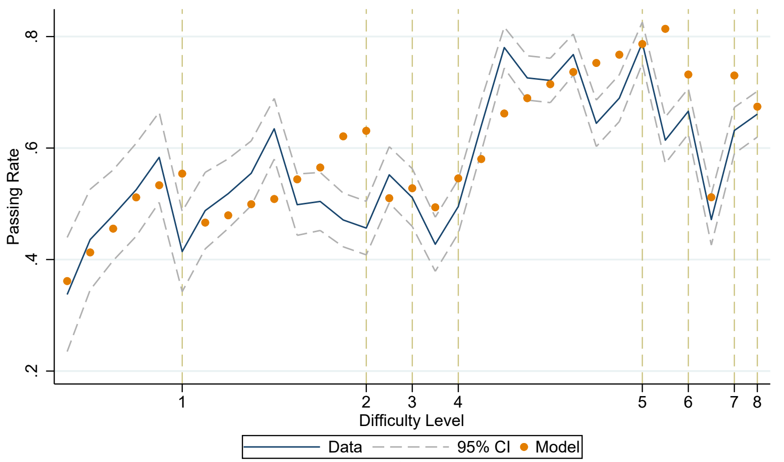
(b) w/o Skill Invariance



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level . Within difficulty levels, tasks are arranged in the order of the children taking them.

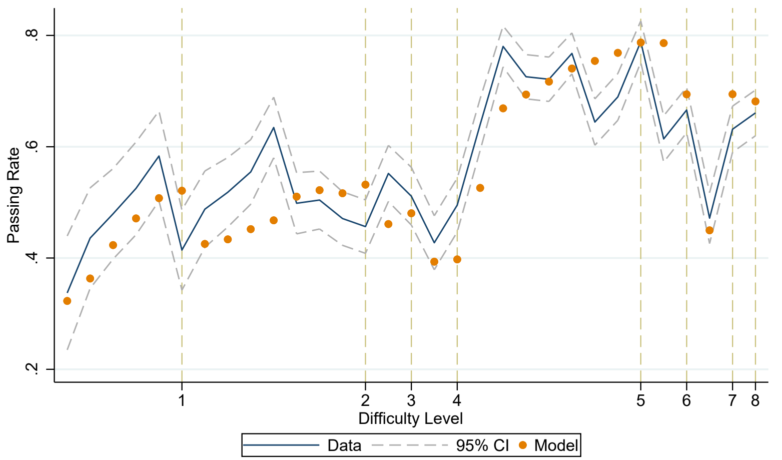
Figure H.11: Fit by Length of Enrollment: Group Enrolled > 1 Month

(a) Skill Invariance



1. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years. 2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.

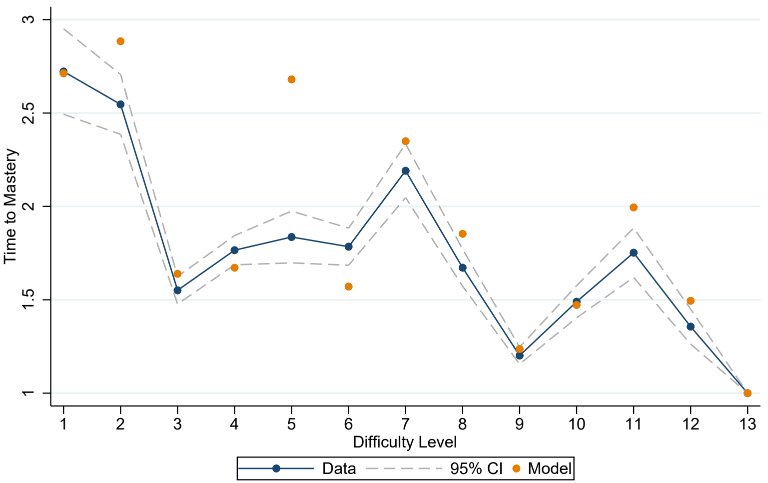
(b) w/o Skill Invariance



1. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years. 2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.

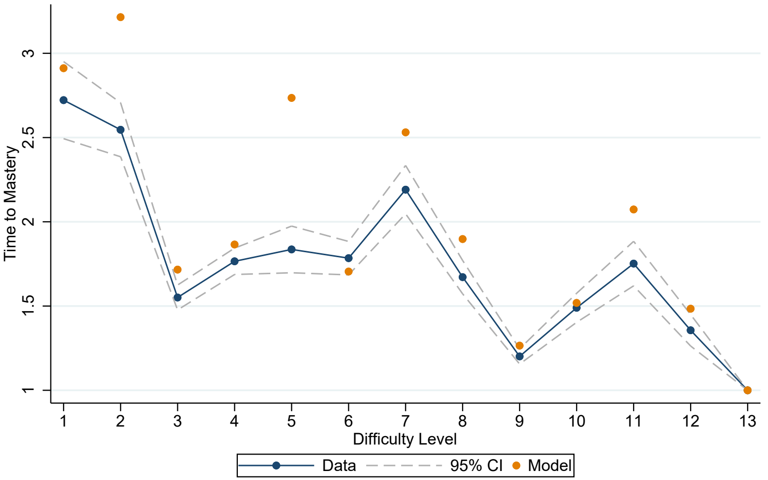
Figure H.12: Fit on Time to Mastery by Level

(a) Skill Invariance



Note: Time to Mastery is defined as the number of tasks a child takes at the previous difficulty level until the first success (inclusive).

(b) w/o Skill Invariance

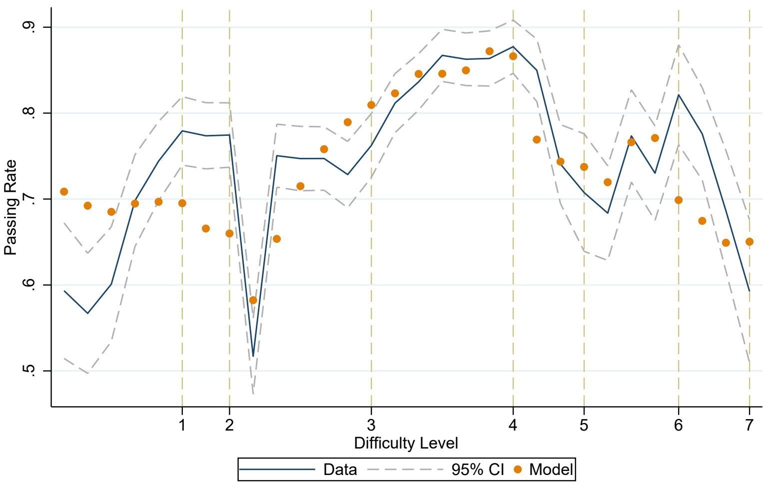


Note: Time to Mastery is defined as the number of tasks a child takes at the previous difficulty level until the first success (inclusive).

H.3 Fine Motor

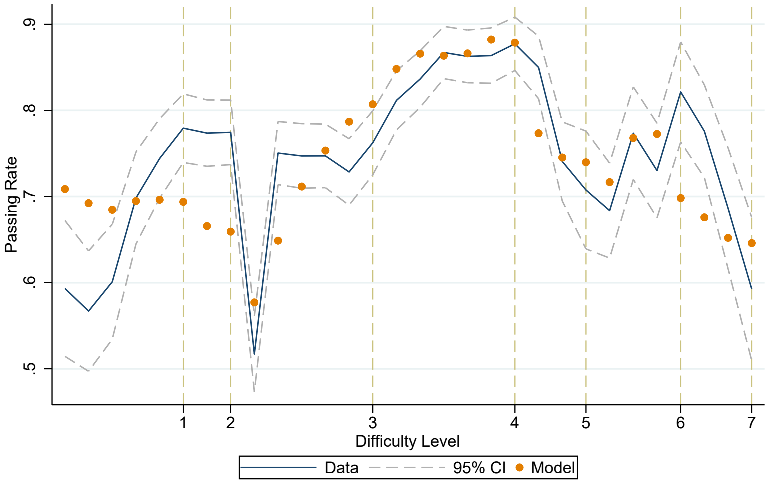
Figure H.13: Fit on All Fine Motor Tasks by Level

(a) Skill Invariance



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

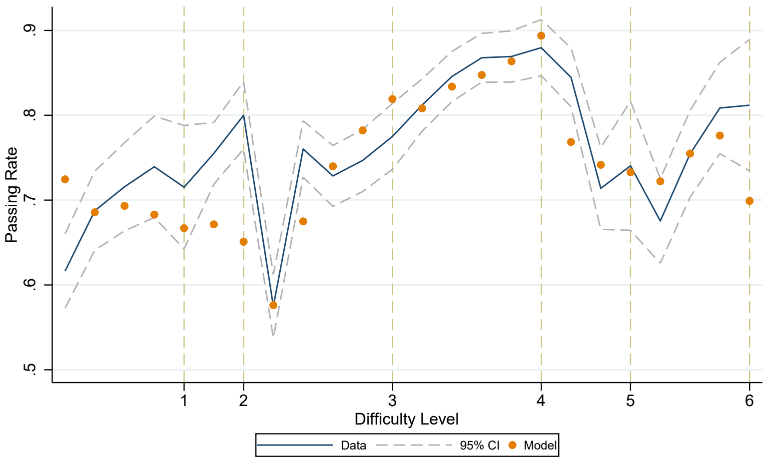
(b) w/o Skill Invariance



Note: The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged by the order of the children taking them.

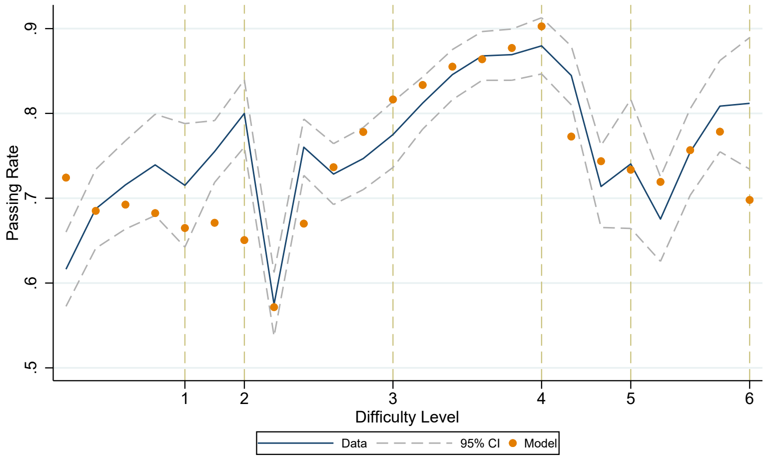
Figure H.14: Fit on the First Five Tasks in Each Level

(a) Skill Invariance



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

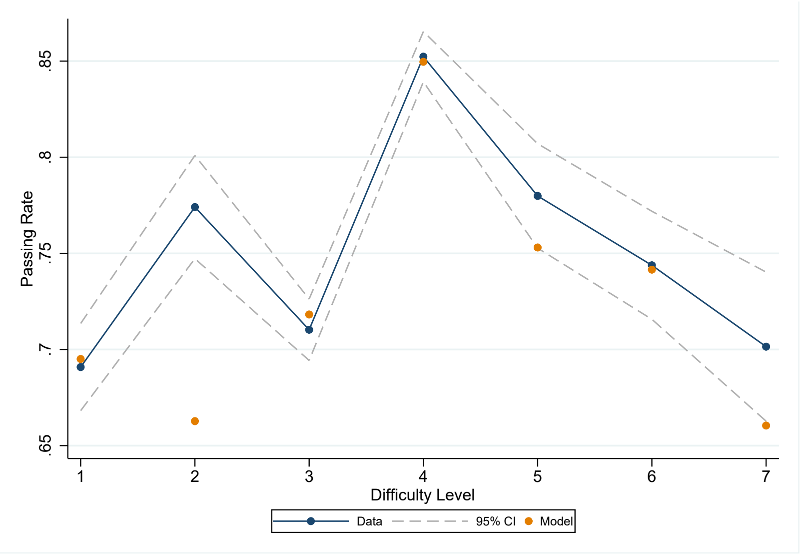
(b) w/o Skill Invariance



Note: The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged by the order of the children taking them.

Figure H.15: Fit on Average Passing Rate by Level

(a) Skill Invariance



(b) w/o Skill Invariance

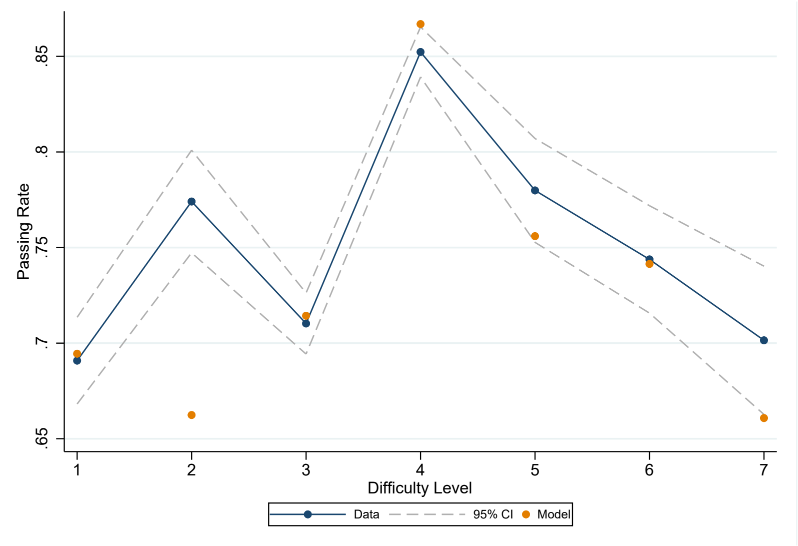
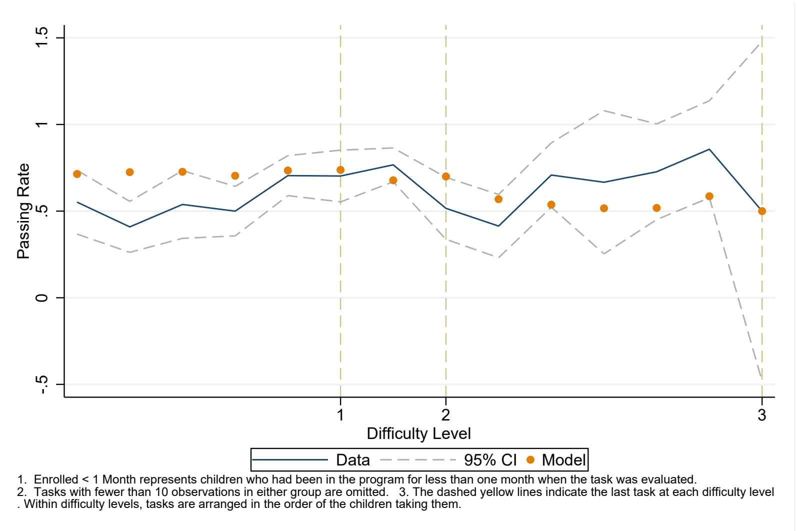


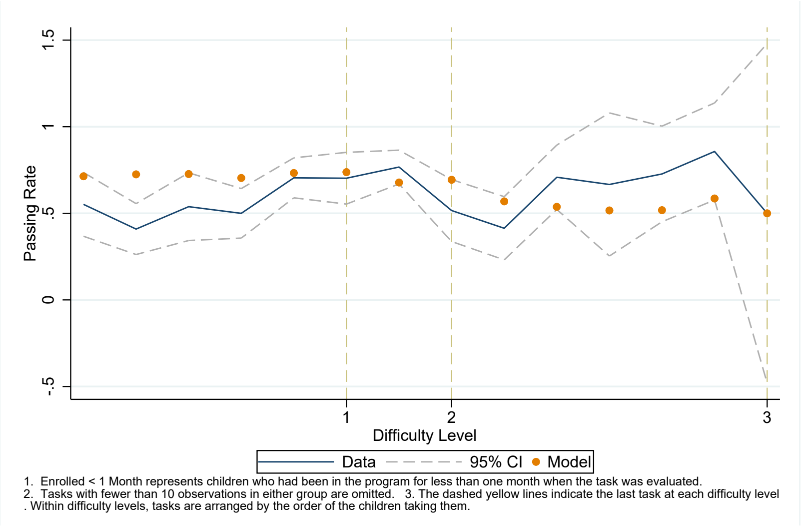
Figure H.16: Fit by Length of Enrollment: Newly Enrolled Group

(a) Skill Invariance



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level . Within difficulty levels, tasks are arranged in the order of the children taking them.

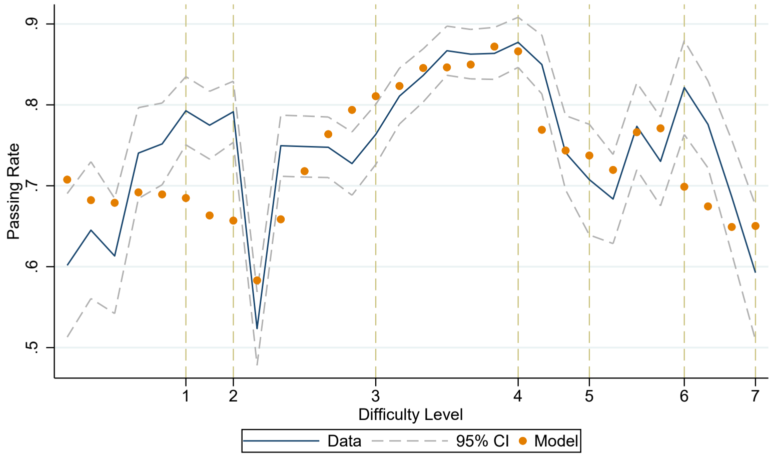
(b) w/o Skill Invariance



1. Enrolled < 1 Month represents children who had been in the program for less than one month when the task was evaluated.
2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level . Within difficulty levels, tasks are arranged by the order of the children taking them.

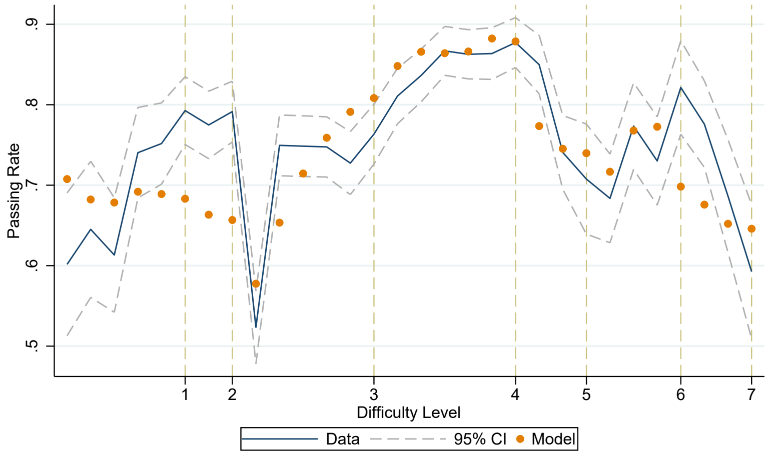
Figure H.17: Fit by Length of Enrollment: Group Enrolled > 1 Month

(a) Skill Invariance



1. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years. 2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them.

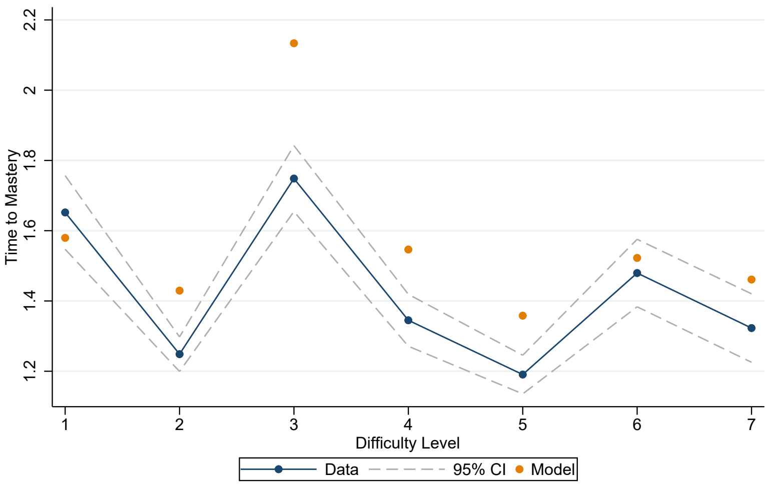
(b) w/o Skill Invariance



1. Enrolled > 1 Month represents children who had been in the program for more than one month when the task was evaluated, who continued to stay in the program for 2 years. 2. Tasks with fewer than 10 observations in either group are omitted. 3. The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged by the order of the children taking them.

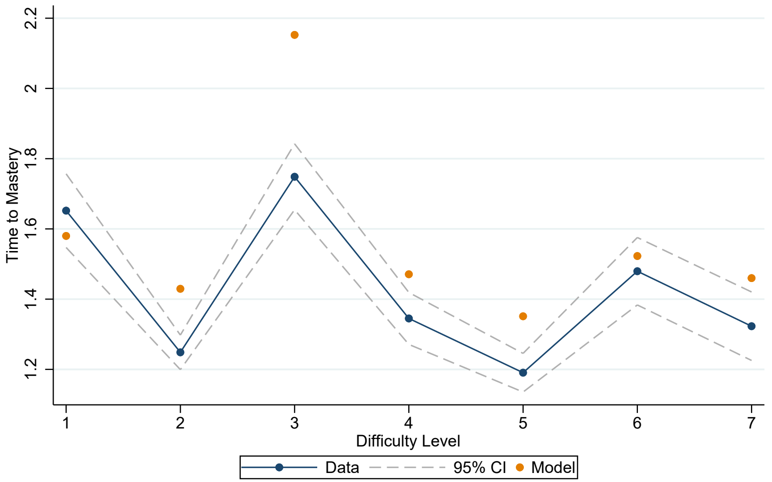
Figure H.18: Fit on Time to Mastery by Level

(a) Skill Invariance



Note: Time to Mastery is defined as the number of tasks a child takes at the previous difficulty level until the first success (inclusive).

(b) w/o Skill Invariance



Note: Time to Mastery is defined as the number of tasks a child takes at the previous difficulty level until the first success (inclusive).

H.4 Language Skill Moment Fit Summary

- Overall, our estimates fit the moments very well. The model w/o skill invariance has better fit.
- χ^2 test results are reported in column 1 of Table H.3.
- We also examine the model of fit by the following summary measure:

$$R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$$

where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment i. Values of R are reported in column 2 of Table H.3.

Table H.3: Goodness of Fit Summary (Language)

	$\chi^2(895)$	R
With skill Invariance	32.71	133.05
Without skill Invariance	21.27	121.39

1. $R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$, where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment.
2. For both models, we cannot reject the model at the 0.0001 level.

H.5 Cognitive Skill Moment Fit Summary

- Overall, our estimates fit the moments very well. The model w/o skill invariance has better fit.
- χ^2 test results are reported in column 1 of Table H.4.

- We also examine the model of fit by the following summary measure:

$$R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$$

where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment i. Values of R are reported in column 2 of Table H.4.

Table H.4: Goodness of Fit Summary (Cognitive)

	$\chi^2(696)$	R
With skill Invariance	16.84	93.5
Without skill Invariance	14.31	81.63

1. $R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$, where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment.
2. For both models, we cannot reject the model at the 0.0001 level.

H.6 Fine Motor Skill Moment Fit Summary

- Similar to the cognitive results, our estimates fit the moments very well. The model w/o skill invariance has better fit.

- χ^2 test results are reported in column 1 of Table H.5.

- We also examine the model of fit by the following equation

$$R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$$

where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment i. Values of R are reported in column 2 of Table H.5.

Table H.5: Goodness of Fit Summary (Fine Motor)

	$\chi^2(339)$	R
With skill Invariance	5.59	29.09
Without skill Invariance	5.23	24.73

1. $R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$, where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment.

2. For both models, we cannot reject the model at the 0.0001 level.

I Point Estimates

I.1 Language Skills

I.1.1 Model with Skill Invariance

Table I.1: Determinants of Initial Conditions and Maturation Effects

Initial Conditions $\mu_0(Z)$		
Male	Point Estimate	–0.001
	Standard Error	(0.011)
Father's years of education	Point Estimate	0.061
	Standard Error	(0.001)
Mother's years of education	Point Estimate	0.056
	Standard Error	(0.001)
Grandmother's years of education	Point Estimate	0.050
	Standard Error	(1.00e-4)
Monthly age of enrollment to the program	Point Estimate	–0.003
	Standard Error	(0.004)
Constant	Point Estimate	–3.070
	Standard Error	(0.680)
Variance of Shock	Point Estimate	3.033
	Standard Error	(0.100)
Maturation Effects		
Child's Age: Month	Point Estimate	7.792e-5
	Standard Error	(2.238e-4)
Child's Age: Week	Point Estimate	–0.013
	Standard Error	(0.009)

1. Standard errors are calculated by 1000 iteration bootstrap.
2. If the child is 8 months and 3 three weeks old, the values for Child's Age: Month, and Child's Age: Week are 8 and 3, respectively.

Table I.2: Minimum Latent Skills Requirement ($\bar{K}$) for Each Level

Level 3	Point Estimate	1.307	Level 8	Point Estimate	4.513
	Standard Error	(0.053)		Standard Error	(0.393)
Level 4	Point Estimate	2.225	Level 9	Point Estimate	5.527
	Standard Error	(0.324)		Standard Error	(0.405)
Level 5	Point Estimate	2.792	Level 10	Point Estimate	10.476
	Standard Error	(0.366)		Standard Error	(0.783)
Level 6	Point Estimate	3.155	Level 11	Point Estimate	11.844
	Standard Error	(0.373)		Standard Error	(0.782)
Level 7	Point Estimate	3.448			
	Standard Error	(0.378)			

1. Standard errors are calculated by 1000 iteration bootstrap.
2. All children started from level 2 or above upon enrolling.
3. Level 2 value is normalized to 1.

Table I.3: Variances of Task Shocks ($\sigma^2_{\varepsilon(\ell)}$) at Each Level

Level 3	Point Estimate	0.993	Level 8	Point Estimate	21.882
	Standard Error	(0.004)		Standard Error	(19.434)
Level 4	Point Estimate	1.502	Level 9	Point Estimate	0.999
	Standard Error	(0.079)		Standard Error	(0.004)
Level 5	Point Estimate	0.990	Level 10	Point Estimate	0.997
	Standard Error	(0.004)		Standard Error	(0.004)
Level 6	Point Estimate	5.895	Level 11	Point Estimate	57.845
	Standard Error	(1.359)		Standard Error	(34.169)
Level 7	Point Estimate	0.998			
	Standard Error	(0.004)			

1. Standard errors are calculated by 1000 iteration bootstrap.
2. All children started from level 2 or above upon enrolling.
3. Level 2 variance is normalized to 1.

Table I.4: Learning Component ($\delta_k \eta(X)$)

	η		δ_k			
Interaction quality:	Point Estimate	0.839	Level 3	3.053	Level 9	5.990
Home visitor and caregiver	Standard Error	(0.444)		(0.005)		(0.007)
Interaction quality:	Point Estimate	0.268	Level	3.049	Level 10	2.998
Home visitor and child	Standard Error	(0.188)		(0.006)		(0.002)
Teaching ability	Point Estimate	0.434	Level 5	0.033	Level 11	2.999
	Standard Error	(0.096)		(0.015)		(0.002)
Grandmother rearing	Point Estimate	–0.230	Level 6	3.111		
	Standard Error	(0.043)		(0.012)		
Male	Point Estimate	–0.073	Level 7	6.000		
	Standard Error	(0.014)		(0.006)		
Constant	Point Estimate	–6.722	Level 8	5.958		
	Standard Error	(0.768)		(0.012)		
Variance of learning ability shock (σ^2_ω)	Point Estimate	0.933				
	Standard Error	(0.008)				

1. Standard errors are calculated by 1000 iteration bootstrap.

2. All children started from level 2 or above upon enrolling. δ_2 at Level 2 is normalized to 1.

3. In the model, we consider the following setting $\delta_k \bar{\eta} = \delta_k(X'\beta_\eta + \omega_i)$. The coefficients of X in η are the same across all levels, and δ_k is level specific.

I.1.2 Model without Skill Invariance

Table I.5: Determinants of Initial Conditions and Maturation Effects

Initial Conditions $\mu_0(Z)$		
Male	Point Estimate	–0.001
	Standard Error	(1.10e-4)
Father's years of education	Point Estimate	0.063
	Standard Error	(0.001)
Mother's years of education	Point Estimate	0.056
	Standard Error	(0.001)
Grandmother's years of education	Point Estimate	0.050
	Standard Error	(2.00e-4)
Monthly age of enrollment to the program	Point Estimate	–0.003
	Standard Error	(3.50e-4)
Constant	Point Estimate	–2.254
	Standard Error	(0.598)
Variance of Shock	Point Estimate	3.043
	Standard Error	(0.105)
Maturation Effects		
Child's Age: Month	Point Estimate	0.016
	Standard Error	(0.005)
Child's Age: Week	Point Estimate	0.002
	Standard Error	(0.001)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. If the child is 8 months and 3 three weeks old, the values for Child's Age: Month, and Child's Age: Week are 8 and 3, respectively.

Table I.6: Minimum Latent Skills Requirement ($\bar{K}$) for Each Level

Level 3	Point Estimate	1.295	Level 8	Point Estimate	4.598
	Standard Error	(0.054)		Standard Error	(0.277)
Level 4	Point Estimate	1.846	Level 9	Point Estimate	5.743
	Standard Error	(0.138)		Standard Error	(0.312)
Level 5	Point Estimate	2.545	Level 10	Point Estimate	13.779
	Standard Error	(0.216)		Standard Error	(1.248)
Level 6	Point Estimate	3.219	Level 11	Point Estimate	15.795
	Standard Error	(0.243)		Standard Error	(1.273)
Level 7	Point Estimate	3.693			
	Standard Error	(0.258)			

1. Standard errors are calculated by 1000 iteration bootstrap.

2. All children started from level 2 or above upon enrolling.

3. Level 2 value is normalized to 1.

Table I.7: Variances of Task Shocks ($\sigma^2_{\varepsilon(\ell)}$) at Each Level

Level 3	Point Estimate	0.992	Level 8	Point Estimate	88.239
	Standard Error	(0.001)		Standard Error	(30.634)
Level 4	Point Estimate	1.384	Level 9	Point Estimate	0.999
	Standard Error	(0.060)		Standard Error	(5.30e-5)
Level 5	Point Estimate	0.990	Level 10	Point Estimate	0.995
	Standard Error	(0.001)		Standard Error	(0.001)
Level 6	Point Estimate	6.293	Level 11	Point Estimate	138.791
	Standard Error	(2.252)		Standard Error	(85.175)
Level 7	Point Estimate	0.999			
	Standard Error	(1.00e-4)			

1. Standard errors are calculated by 1000 iteration bootstrap.

2. All children started from level 2 or above upon enrolling. Level 2 variance is normalized to 1.

Table I.8: Learning Component ($\delta_k \eta(X)$)

η			δ_k			
Interaction quality:	Point Estimate	0.664	Level 3	3.053	Level 10	2.998
Home visitor and caregiver	Standard Error	(0.149)		(0.005)		(2.00e-4)
Interaction quality:	Point Estimate	0.245	Level 4	3.048	Level 11	2.999
Home visitor and child	Standard Error	(0.139)		(0.005)		(1.00e-4)
Teaching ability	Point Estimate	0.364	Level 5	0.023		
	Standard Error	(0.051)		(0.066)		
Grandmother rearing	Point Estimate	–0.260	Level 6	3.108		
	Standard Error	(0.027)		(0.013)		
Male	Point Estimate	–0.073	Level 7	6.000		
	Standard Error	(0.008)		(0.039)		
Constant	Point Estimate	–6.169	Level 8	5.986		
	Standard Error	(1.462)		(0.028)		
Variance of learning ability	Point Estimate	0.928	Level 9	5.970		
shock (σ^2_ω)	Standard Error	(0.007)		(0.015)		

1. Standard errors are calculated by 1000 iteration bootstrap.

2. All children started from level 2 or above upon enrolling. δ_2 at Level 2 is normalized to 1.

3. In the model, we consider the following setting $\delta_k \bar{\eta} = \delta_k(X'\beta_\eta + \omega_i)$. The coefficients of X in η are the same across all levels, and δ_k is level specific.

Table I.9: Transformation Functions

		$\gamma_{k,\ell}$			$\gamma_{k,\ell}$
Level 2	Point Estimate	0.460	Level 7	Point Estimate	1.125
	Standard Error	(0.071)		Standard Error	(0.293)
Level 3	Point Estimate	0.901	Level 8	Point Estimate	0.562
	Standard Error	(0.134)		Standard Error	(0.153)
Level 4	Point Estimate	0.645	Level 9	Point Estimate	1.113
	Standard Error	(0.079)		Standard Error	(0.336)
Level 5	Point Estimate	0.660	Level 10	Point Estimate	1.006
	Standard Error	(0.111)		Standard Error	(0.160)
Level 6	Point Estimate	1.522	Level 11	Point Estimate	1.223
	Standard Error	(0.232)		Standard Error	(0.364)

1. Standard errors are calculated by 1000 iteration bootstrap.
2. Our language task data starts from level 2.

I.2 Cognitive Skills

I.2.1 Model with Skill Invariance

Table I.10: Determinants of Initial Conditions and Maturation Effects

Initial Conditions $\mu_0(Z)$		
Male	Point Estimate	–0.552
	Standard Error	(0.139)
Father's years of education	Point Estimate	0.058
	Standard Error	(0.001)
Mother's years of education	Point Estimate	0.052
	Standard Error	(2.30e-4)
Grandmother's years of education	Point Estimate	0.050
	Standard Error	(1.00e-4)
Monthly age of enrollment to the program	Point Estimate	0.471
	Standard Error	(0.052)
Constant	Point Estimate	–7.921
	Standard Error	(0.749)
Variance of Shock	Point Estimate	2.050
	Standard Error	(0.008)
Maturation Effects		
Child's Age: Month	Point Estimate	-3.450e-4
	Standard Error	(0.001)
Child's Age: Week	Point Estimate	1.529e-5
	Standard Error	(2.983e-4)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. If the child is 8 months and 3 three weeks old, the values for Child's Age: Month, and Child's Age: Week are 8 and 3, respectively.

Table I.11: Minimum Latent Skills Requirement ($\bar{K}$) for Each Level

Level 2	Point Estimate	2.083	Level 8	Point Estimate	23.317
	Standard Error	(0.202)		Standard Error	(4.058)
Level 3	Point Estimate	3.380	Level 9	Point Estimate	57.728
	Standard Error	(0.318)		Standard Error	(7.227)
Level 4	Point Estimate	5.374	Level 10	Point Estimate	89.071
	Standard Error	(0.574)		Standard Error	(8.336)
Level 5	Point Estimate	5.712	Level 11	Point Estimate	90.586
	Standard Error	(0.582)		Standard Error	(8.311)
Level 6	Point Estimate	7.179	Level 12	Point Estimate	91.901
	Standard Error	(0.637)		Standard Error	(8.333)
Level 7	Point Estimate	21.729	Level 13	Point Estimate	5856.449
	Standard Error	(4.086)		Standard Error	(1023.389)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Level 1 value is normalized to 1.

Table I.12: Variances of Task Shocks ($\sigma^2_{\varepsilon(\ell)}$) at Each Level

Level 2	Point Estimate	1.246	Level 8	Point Estimate	64.556
	Standard Error	(0.083)		Standard Error	(75.766)
Level 3	Point Estimate	1.206	Level 9	Point Estimate	0.999
	Standard Error	(0.075)		Standard Error	(0.004)
Level 4	Point Estimate	0.977	Level 10	Point Estimate	175.186
	Standard Error	(0.015)		Standard Error	(372.405)
Level 5	Point Estimate	1.000	Level 11	Point Estimate	1.000
	Standard Error	(0.004)		Standard Error	(0.004)
Level 6	Point Estimate	1.128	Level 12	Point Estimate	265.272
	Standard Error	(0.022)		Standard Error	(850.998)
Level 7	Point Estimate	1.110			
	Standard Error	(0.017)			

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Level 1 and 13 variances are normalized to 1.

Table I.13: Learning Component ($\delta_k \eta(X)$)

η			δ_k			
Interaction quality:	Point Estimate	0.954	Level 2	0.872	Level 8	2.000
Home visitor and caregiver	Standard Error	(0.555)		(0.226)		(0.001)
Interaction quality:	Point Estimate	0.002	Level 5	2.218	Level 10	1.998
Home visitor and child	Standard Error	(0.012)		(0.052)		(0.001)
Teaching ability	Point Estimate	0.026	Level 6	2.173	Level 11	3.997
	Standard Error	(0.016)		(0.025)		(0.027)
Grandmother rearing	Point Estimate	−0.037	Level 7	3.999	Level 12	2.000
	Standard Error	(0.012)		(0.008)		(0.001)
Male	Point Estimate	0.002				
	Standard Error	(0.012)				
Constant	Point Estimate	−3.747				
	Standard Error	(1.201)				
Variance of learning ability shock (σ_ω^2)	Point Estimate	0.986				
	Standard Error	(0.004)				

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Since the number of tasks at level 1, 3, 4, 9, and 13 are less than 3, we normalize the values of δ_k to 1.

3. In the model, we consider the following setting $\delta_k \bar{\eta} = \delta_k(X'\beta_\eta + \omega_i)$. The coefficients of X in η are the same across all levels, and δ_k is level specific.

I.2.2 Model without Skill Invariance

Table I.14: Determinants of Initial Conditions and Maturation Effects

Initial Conditions $\mu_0(Z)$		
Male	Point Estimate	–0.580
	Standard Error	(0.100)
Father's years of education	Point Estimate	0.058
	Standard Error	(0.001)
Mother's years of education	Point Estimate	0.052
	Standard Error	(1.567e-4)
Grandmother's years of education	Point Estimate	0.050
	Standard Error	(6.210e-7)
Monthly age of enrollment to the program	Point Estimate	0.437
	Standard Error	(0.034)
Constant	Point Estimate	–8.067
	Standard Error	(0.560)
Variance of Shock	Point Estimate	2.051
	Standard Error	(0.008)
Maturation Effects		
Child's Age: Month	Point Estimate	-3.622e-4
	Standard Error	(2.306e-5)
Child's Age: Week	Point Estimate	9.386e-5
	Standard Error	(2.430e-5)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. If the child is 8 months and 3 three weeks old, the values for Child's Age: Month, and Child's Age: Week are 8 and 3, respectively.

Table I.15: Minimum Latent Skills Requirement ($\bar{K}$) for Each Level

Level 2	Point Estimate	2.127	Level 8	Point Estimate	21.662
	Standard Error	(0.156)		Standard Error	(2.959)
Level 3	Point Estimate	3.468	Level 9	Point Estimate	56.722
	Standard Error	(0.258)		Standard Error	(5.510)
Level 4	Point Estimate	5.472	Level 10	Point Estimate	89.275
	Standard Error	(0.508)		Standard Error	(7.329)
Level 5	Point Estimate	5.813	Level 11	Point Estimate	90.812
	Standard Error	(0.514)		Standard Error	(7.340)
Level 6	Point Estimate	7.271	Level 12	Point Estimate	92.136
	Standard Error	(0.590)		Standard Error	(7.344)
Level 7	Point Estimate	19.361	Level 13	Point Estimate	5817.490
	Standard Error	(2.948)		Standard Error	(791.794)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Level 1 value is normalized to 1.

Table I.16: Variances of Task Shocks ($\sigma^2_{\varepsilon(\ell)}$) at Each Level

Level 2	Point Estimate	1.238	Level 8	Point Estimate	61.185
	Standard Error	(0.055)		Standard Error	(39.474)
Level 3	Point Estimate	1.236	Level 9	Point Estimate	0.999
	Standard Error	(0.044)		Standard Error	(7.00e-5)
Level 4	Point Estimate	0.978	Level 10	Point Estimate	195.013
	Standard Error	(0.003)		Standard Error	(384.660)
Level 5	Point Estimate	1.000	Level 11	Point Estimate	1.000
	Standard Error	(1.00e-5)		Standard Error	(3.00e-5)
Level 6	Point Estimate	1.128	Level 12	Point Estimate	335.661
	Standard Error	(0.018)		Standard Error	(700.474)
Level 7	Point Estimate	1.110			
	Standard Error	(0.017)			

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Level 1 and 13 variances are normalized to 1.

Table I.17: Learning Component ($\delta_k\eta(X)$)

η			δ_k			
Interaction quality:	Point Estimate	0.938	Level 2	0.863	Level 8	2.000
Home visitor and caregiver	Standard Error	(0.421)		(0.168)		(1.00e-6)
Interaction quality:	Point Estimate	0.002	Level 5	2.204	Level 10	1.998
Home visitor and child	Standard Error	(2.00e-4)		(0.072)		(2.00e-4)
Teaching ability	Point Estimate	0.026	Level 6	2.181	Level 11	3.997
	Standard Error	(0.004)		(0.028)		(0.015)
Grandmother rearing	Point Estimate	−0.037	Level 7	4.000	Level 12	2.000
	Standard Error	(0.005)		(0.002)		(1.00e-4)
Male	Point Estimate	0.002				
	Standard Error	(2.00e-4)				
Constant	Point Estimate	−3.763				
	Standard Error	(1.954)				
Variance of learning ability	Point Estimate	0.986				
shock (σ^2_ω)	Standard Error	(0.002)				

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Since the number of tasks at level 1, 3, 4, 9, and 13 are less than 3, we normalize the values of δ_k to 1.

3. In the model, we consider the following setting $\delta_k\bar{\eta} = \delta_k(X'\beta_\eta + \omega_i)$. The coefficients of X in η are the same across all levels and δ_k is level specific.

Table I.18: Transformation Functions

		$\gamma_{k,\ell}$			$\gamma_{k,\ell}$
Level 1	Point Estimate	0.800	Level 8	Point Estimate	1.893
	Standard Error	(0.101)		Standard Error	(0.434)
Level 2	Point Estimate	0.929	Level 9	Point Estimate	0.744
	Standard Error	(0.655)		Standard Error	(0.138)
Level 3	Point Estimate	0.936	Level 10	Point Estimate	2.068
	Standard Error	(0.651)		Standard Error	(0.306)
Level 4	Point Estimate	0.621	Level 11	Point Estimate	2.292
	Standard Error	(1.007)		Standard Error	(0.391)
Level 5	Point Estimate	2.235	Level 12	Point Estimate	5.614
	Standard Error	(0.625)		Standard Error	(1.218)
Level 6	Point Estimate	0.317	Level 13	Point Estimate	1.420
	Standard Error	(0.163)		Standard Error	(0.202)
Level 7	Point Estimate	0.791			
	Standard Error	(0.347)			

1. Standard errors are calculated by 1000 iteration bootstrap.

I.3 Fine Motor Skills

I.3.1 Model with Skill Invariance

Table I.19: Determinants of Initial Conditions and Maturation Effects

Initial Conditions $\mu_0(Z)$		
Male	Point Estimate	0.067
	Standard Error	(0.014)
Father's years of education	Point Estimate	0.035
	Standard Error	(0.002)
Mother's years of education	Point Estimate	0.049
	Standard Error	(1.40e-4)
Grandmother's years of education	Point Estimate	0.072
	Standard Error	(0.002)
Monthly age of enrollment to the program	Point Estimate	0.037
	Standard Error	(0.014)
Constant	Point Estimate	0.206
	Standard Error	(0.047)
Variance of Shock	Point Estimate	2.087
	Standard Error	(0.011)
Maturation Effects		
Child's Age: Month	Point Estimate	-9.414e-5
	Standard Error	(2.133e-4)
Child's Age: Week	Point Estimate	-1.807e-4
	Standard Error	(2.740e-4)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. If the child is 8 months and 3 three weeks old, the values for Child's Age: Month, and Age Child's week are 8 and 3, respectively.

Table I.20: Minimum Latent Skills Requirement ($\bar{K}$) for Each Level

Level 2	Point Estimate	1.004
	Standard Error	(0.011)
Level 3	Point Estimate	3.238
	Standard Error	(0.568)
Level 4	Point Estimate	3.304
	Standard Error	(0.568)
Level 5	Point Estimate	4.993
	Standard Error	(0.613)
Level 6	Point Estimate	21.789
	Standard Error	(3.972)
Level 7	Point Estimate	53.818
	Standard Error	(7.751)

1. Standard errors are calculated by 1000 iteration bootstrap.
2. Level 1 value is normalized to 1.

Table I.21: Variances of Task Shocks ($\sigma^2_{\varepsilon(\ell)}$) at Each Level

Level 2	Point Estimate	2.033
	Standard Error	(0.332)
Level 3	Point Estimate	1.000
	Standard Error	(0.004)
Level 4	Point Estimate	1.000
	Standard Error	(0.004)
Level 5	Point Estimate	28.002
	Standard Error	(18.843)
Level 6	Point Estimate	16.359
	Standard Error	(27.517)
Level 7	Point Estimate	172.202
	Standard Error	(281.399)

1. Standard errors are calculated by 1000 iteration bootstrap.
2. Level 1 variance is normalized to 1.

Table I.22: Learning Component ($\delta_k \eta(X)$)

η			δ_k	
Interaction quality: Home visitor and caregiver	Point Estimate	0.026	Level 3	2.619
	Standard Error	(0.023)		(0.126)
Interaction quality: Home visitor and child	Point Estimate	0.039	Level 4	4.000
	Standard Error	(0.018)		(0.002)
Teaching ability	Point Estimate	1.50e-4	Level 5	2.000
	Standard Error	(0.012)		(0.004)
Grandmother rearing	Point Estimate	−0.007	Level 6	3.899
	Standard Error	(0.011)		(0.067)
Male	Point Estimate	0.164	Level 7	3.318
	Standard Error	(0.033)		(0.100)
Constant	Point Estimate	−5.849		
	Standard Error	(0.603)		
Variance of learning ability shock (σ^2_ω)	Point Estimate	1.015		
	Standard Error	(0.004)		

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Since the number of tasks at level 1 and 2 are less than 3, we normalize the values of δ_k to 1.

3. In the model, we consider the following setting $\delta_k \bar{\eta} = \delta_k(X'\beta_\eta + \omega_i)$. The coefficients of X in η are the same across all levels, and δ_k is level specific.

I.3.2 Model without Skill Invariance

Table I.23: Determinants of Initial Conditions and Maturation Effects

Initial Conditions $\mu_0(Z)$		
Male	Point Estimate	0.067
	Standard Error	(0.021)
Father's years of education	Point Estimate	0.035
	Standard Error	(0.004)
Mother's years of education	Point Estimate	0.049
	Standard Error	(2.90e-4)
Grandmother's years of education	Point Estimate	0.072
	Standard Error	(0.004)
Monthly age of enrollment to the program	Point Estimate	0.037
	Standard Error	(0.013)
Constant	Point Estimate	0.206
	Standard Error	(0.060)
Variance of Shock	Point Estimate	2.087
	Standard Error	(0.024)
Maturation Effects		
Child's Age: Month	Point Estimate	-1.820e-4
	Standard Error	(0.006)
Child's Age: Week	Point Estimate	-4.352e-5
	Standard Error	(0.012)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. If the child is 8 months and 3 three weeks old, the values for Child's Age: Month, and Child's Age: Week are 8 and 3, respectively.

Table I.24: Minimum Latent Skills Requirement ($\bar{K}$) for Each Level

Level 2	Point Estimate	1.004
	Standard Error	(0.012)
Level 3	Point Estimate	3.238
	Standard Error	(0.729)
Level 4	Point Estimate	3.307
	Standard Error	(0.731)
Level 5	Point Estimate	4.998
	Standard Error	(0.865)
Level 6	Point Estimate	21.814
	Standard Error	(4.912)
Level 7	Point Estimate	57.883
	Standard Error	(15.155)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Level 1 value is normalized to 1.

Table I.25: Variances of Task Shocks ($\sigma^2_{\varepsilon(\ell)}$) at Each Level

Level 2	Point Estimate	2.034
	Standard Error	(0.371)
Level 3	Point Estimate	1.000
	Standard Error	(0.004)
Level 4	Point Estimate	1.000
	Standard Error	(0.004)
Level 5	Point Estimate	28.039
	Standard Error	(28.111)
Level 6	Point Estimate	16.381
	Standard Error	(32.714)
Level 7	Point Estimate	170.883
	Standard Error	(247.758)

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Level 1 variance is normalized to 1.

Table I.26: Learning Component ($\delta_k \eta(X)$)

η			δ_k	
Interaction quality:	Point Estimate	0.026	Level 3	2.619
Home visitor and caregiver	Standard Error	(0.013)		(0.197)
Interaction quality:	Point Estimate	0.039	Level 4	4.000
Home visitor and child	Standard Error	(0.029)		(0.027)
Teaching ability	Point Estimate	1.50e-4	Level 5	2.000
	Standard Error	(0.012)		(0.004)
Grandmother rearing	Point Estimate	–0.007	Level 6	3.932
	Standard Error	(0.011)		(0.156)
Male	Point Estimate	0.164	Level 7	3.319
	Standard Error	(0.050)		(0.182)
Constant	Point Estimate	–5.850		
	Standard Error	(31.366)		
Variance of learning ability shock (σ^2_ω)	Point Estimate	1.015		
	Standard Error	(0.007)		

1. Standard errors are calculated by 1000 iteration bootstrap.

2. Since the number of tasks at level 1, 3, 4, 9, and 13 are less than 3, we normalize the values of δ_k to 1.

3. In the model, we consider the following setting $\delta_k \bar{\eta} = \delta_k(X'\beta_\eta + \omega_i)$. The coefficients of X in η are the same across all levels, and δ_k is level specific.

Table I.27: Transformation Function

		$\gamma_{k,\ell}$
Level 1	Point Estimate	1.365
	Standard Error	(0.307)
Level 2	Point Estimate	1.005
	Standard Error	(0.520)
Level 3	Point Estimate	0.963
	Standard Error	(0.252)
Level 4	Point Estimate	1.446
	Standard Error	(0.507)
Level 5	Point Estimate	0.798
	Standard Error	(0.238)
Level 6	Point Estimate	0.748
	Standard Error	(0.223)
Level 7	Point Estimate	0.955
	Standard Error	(0.243)

1. Standard errors are calculated by 1000 iteration bootstrap.

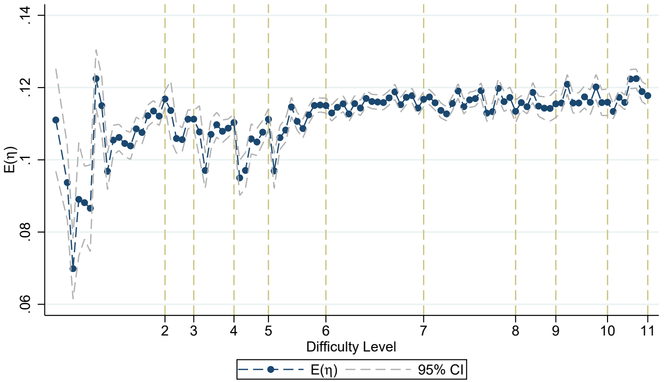
Table I.28: Skill Invariance Hypothesis Tests by Levels

	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 2				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		44.051	0.000		71.398	0.000		2.827	0.830

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 skill invariance assumption is valid across all the levels.
4. Our data for language tasks starts from level 2.

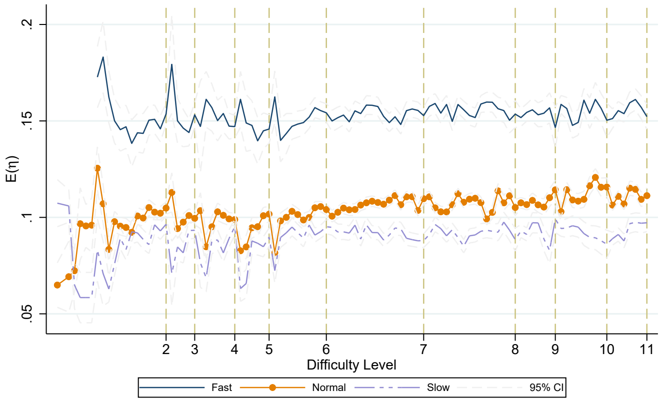
J Learning Component

Figure J.1: Learning Component $E(\eta(X))$ of Language Tasks by Level



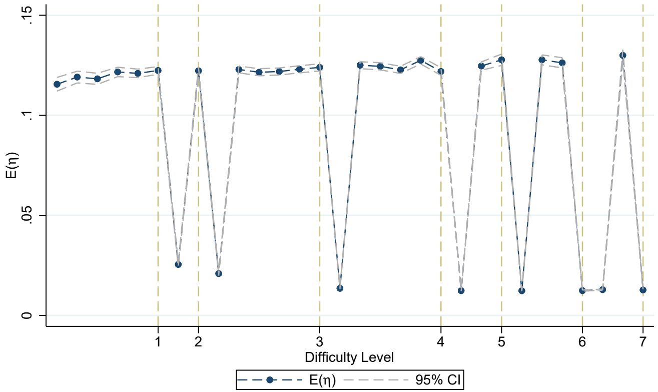
Note: The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the curriculum design.
All the children started from level 2 or above upon enrolling.

Figure J.2: Learning Component $E(\eta(X))$ of Language Tasks by Level and Ability Group



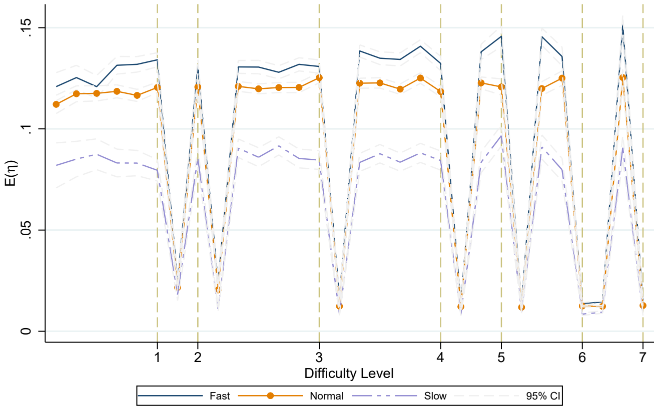
1. Fast group: the child can pass the first task at over 80% of the difficulty levels, and the average pass rate at that level is greater than 80%. Normal group: the child doesn't pass the first task, and the pass rate is greater than 50%; or the child passes the first task, and the pass rate is between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

Figure J.3: Learning Component $E(\eta(X))$ of Fine Motor 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 J.4: Learning Component $E(\eta(X))$ of Fine Motor 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.

Table J.1: The Comparison of Interaction Components on Cognitive Skill Tasks' Learning Components by Family Educational Background

Mean	Teaching Ability	Interaction Quality Home Visitor and Caregiver	Interaction Quality Home Visitor and Child
Father's Years of Education >9 Mean	1.0017	1.2745	1.0001
Father's Years of Education ≤9 Mean	1.0010	1.1641	1.0000
p-value	**(0.0012)**	**(0.0000)**	**(0.0190)**
Mother's Years of Education >9 Mean	1.0030	1.2996	1.0004
Mother's Years of Education ≤9 Mean	1.0009	1.1701	1.0000
p-value	**(0.0000)**	**(0.0000)**	**(0.0000)**
Grandmother's Years of Education≥3 Mean	1.0015	1.2273	1.0002
Grandmother's Years of Education <3 Mean	1.0009	1.1569	1.0000
p-value	**(0.0002)**	**(0.0000)**	**(0.0000)**

1. About 40% of children's grandmothers have more than 3 years of formal education.
2. *p*-values are presented in the parentheses.

Table J.2: The Comparison of Interaction Components by Cognitive Skill Ability

Mean	Teaching Ability	Interaction Quality Home Visitor and Caregiver	Interaction Quality Home Visitor and Child
Slow Mean	0.9955	0.8262	0.9988
Fast Mean	1.0042	1.4681	1.0009
p-value	**(0.0000)**	**(0.0000)**	**(0.0000)**
Normal Mean	1.0006	1.1202	0.9999
Fast Mean	1.0042	1.4681	1.0009
p-value	**(0.0000)**	**(0.0000)**	**(0.0000)**
Slow Mean	0.9955	0.8262	0.9988
Normal Mean	1.0006	1.1202	0.9999
p-value	**(0.0000)**	**(0.0000)**	**(0.0000)**

1. Ability Definition: Fast group: child passes the first tasks for more than 80% difficulty levels, and the average passing rate is more than 80% of all the tasks. Normal group: Child passes the first tasks less than 80% of difficulty levels, and the average passing rate is between 50% and 80%. Slow group: the average passing rate is less than 50%.
2. *p*-values are presented in the parentheses.

Table J.3: The Comparison of Interaction Components by Cognitive Skill Ability and Age of Enrollment

Ability	Age of Enrollment	Teaching Ability	Interaction Quality Home Visitor and Caregiver	Interaction Quality Home Visitor and Child
Fast	Age enrollment 20–25 Months Mean	1.0045	1.4569	1.0008
	Age enrollment 9–15 Months Mean	1.0053	1.5277	1.0010
	p-value	**(0.0157)**	**(0.0003)**	**(0.0242)**
	Age enrollment 15–20 Months Mean	1.0034	1.4163	1.0008
	Age enrollment 9–15 Months Mean	1.0053	1.5277	1.0010
	p-value	**(0.0000)**	**(0.0000)**	**(0.0017)**
	Age enrollment 20–25 Months Mean	1.0045	1.4569	1.0008
	Age enrollment 15–20 Months Mean	1.0034	1.4163	1.0008
	p-value	**(0.0007)**	**(0.0198)**	(0.6445)
Normal	Age enrollment 20–25 Months Mean	1.0016	1.1802	1.0002
	Age enrollment 9–15 Months Mean	1.0004	1.0611	0.9998
	p-value	**(0.0000)**	**(0.0000)**	**(0.0000)**
	Age enrollment 15–20 Months Mean	1.0002	1.1419	0.9998
	Age enrollment 9–15 Months Mean	1.0004	1.0611	0.9998
	p-value	(0.4940)	**(0.0000)**	(0.4512)
	Age enrollment 20–25 Months Mean	1.0016	1.1802	1.0002
	Age enrollment 15–20 Months Mean	1.0002	1.1419	0.9998
	p-value	**(0.0000)**	**(0.0032)**	**(0.0000)**
Slow	Age enrollment 20–25 Months Mean	0.9965	0.7747	0.9988
	Age enrollment 9–15 Months Mean	0.9974	0.9248	0.9989
	p-value	**(0.0755)**	**(0.0000)**	(0.4359)
	Age enrollment 15–20 Months Mean	0.9927	0.7581	0.9985
	Age enrollment 9–15 Months Mean	0.9974	0.9248	0.9989
	p-value	**(0.0000)**	**(0.0000)**	**(0.0034)**
	Age enrollment 20–25 Months Mean	0.9965	0.7747	0.9988
	Age enrollment 15–20 Months Mean	0.9927	0.7581	0.9985
	p-value	**(0.0000)**	(0.5448)	**(0.0192)**

1. Ability Definition: Fast group: child passes the first tasks for more than 80% difficulty levels, and the average passing rate is more than 80% of all the tasks. Normal group: Child passes the first tasks less than 80% of difficulty levels, and the average passing rate is between 50% and 80%. Slow group: the average passing rate is less than 50%.

2. *p*-values are presented in the parentheses.

Table J.4: The Effects of Interaction Measures on Language Skill Tasks' Learning Components by Family Educational Background

Mean	Teaching Ability	Interaction Quality Home Visitor and Caregiver	Interaction Quality Home Visitor and Child
Father's Educ>9 Mean	1.033	1.158	1.071
Father's Educ≤9 Mean	1.019	1.073	1.049
p-value	**(0.000)**	**(0.000)**	**(0.000)**
Mother's Educ>9 Mean	1.044	1.160	1.102
Mother's Educ≤9 Mean	1.019	1.080	1.046
p-value	**(0.000)**	**(0.000)**	**(0.000)**
Grandmother's Educ>3 Mean	1.029	1.120	1.060
Grandmother's Educ≤3 Mean	1.017	1.069	1.049
p-value	**(0.000)**	**(0.000)**	**(0.000)**

1. About 40% of children's grandmother have more than 3 years of formal education.
2. *p*-values are presented in the parentheses.

Table J.5: The Effects of Interaction Measures on Language Skill Tasks' Learning Components by Language Skill Ability

Mean	Teaching Ability	Interaction Quality Home Visitor and Caregiver	Interaction Quality Home Visitor and Child
Slow Mean	0.978	0.955	0.960
Fast Mean	1.078	1.360	1.181
p-value	**(0.000)**	**(0.000)**	**(0.000)**
Normal Mean	1.010	1.022	1.023
Fast Mean	1.078	1.360	1.181
p-value	**(0.000)**	**(0.000)**	**(0.000)**
Slow Mean	0.978	0.955	0.960
Normal Mean	1.010	1.022	1.023
p-value	**(0.000)**	**(0.000)**	**(0.000)**

1. Ability Definition:

Fast group: child passes the first tasks for more than 80% difficulty levels, and the average passing rate is more than 80% of all the tasks. Normal group: Child passes the first tasks less than 80% of difficulty levels, and the average passing rate is between 50% and 80%. Slow group: the average passing rate is less than 50%.

2. *p*-values are presented in the parentheses.

Table J.6: The Effects of Interaction Measures on Language Skill Tasks' Learning Components by Language Skill Ability and Age of Enrollment

Ability		Teaching	Interaction Quality	Interaction Quality
Fast	Age enrollment 20–25 Months Mean	1.079	1.351	1.186
	Age enrollment 9–15 Months Mean	1.099	1.401	1.179
	p-value	**(0.000)**	**(0.000)**	(0.161)
	Age enrollment 15–20 Months Mean	1.064	1.316	1.188
	Age enrollment 9–15 Months Mean	1.099	1.401	1.179
	p-value	**(0.000)**	**(0.000)**	**(0.027)**
	Age enrollment 20–25 Months Mean	1.079	1.351	1.186
	Age enrollment 15–20 Months Mean	1.064	1.316	1.188
	p-value	**(0.000)**	**(0.000)**	(0.531)
Normal	Age enrollment 20–25 Months Mean	1.020	1.044	1.057
	Age enrollment 9–15 Months Mean	1.013	0.998	1.026
	p-value	**(0.001)**	**(0.000)**	**(0.000)**
	Age enrollment 15–20 Months Mean	1.005	1.047	1.001
	Age enrollment 9–15 Months Mean	1.013	0.998	1.026
	p-value	**(0.000)**	**(0.000)**	**(0.000)**
	Age enrollment 20–25 Months Mean	1.020	1.044	1.057
	Age enrollment 15–20 Months Mean	1.005	1.047	1.001
	p-value	**(0.000)**	(0.528)	**(0.000)**
Slow	Age enrollment 20–25 Months Mean	0.981	0.906	0.971
	Age enrollment 9–15 Months Mean	0.995	0.988	0.952
	p-value	**(0.000)**	**(0.000)**	**(0.001)**
	Age enrollment 15–20 Months Mean	0.945	0.932	0.936
	Age enrollment 9–15 Months Mean	0.995	0.988	0.952
	p-value	**(0.000)**	**(0.000)**	**(0.005)**
	Age enrollment 20–25 Months Mean	0.981	0.906	0.971
	Age enrollment 15–20 Months Mean	0.945	0.932	0.936
	p-value	**(0.000)**	**(0.001)**	**(0.000)**

1. *p*-values are presented in the parentheses.

2. Ability Definition:

Fast group: child passes the first tasks for more than 80% difficulty levels, and the average passing rate is more than 80% of all the tasks. Normal group: Child passes the first tasks less than 80% of difficulty levels, and the average passing rate is between 50% and 80%. Slow group: the average passing rate is less than 50%.

Table J.7: The Effects of Interaction Measures on Fine Motor Skill Tasks' Learning Components by Family Educational Background

Mean	Teaching Ability	Interaction Quality Home Visitor and Caregiver	Interaction Quality Home Visitor and Child
Father's Educ>9 Mean	1.000	1.003	1.007
Father's Educ≤9 Mean	1.000	1.000	1.002
p-value	**(0.000)**	**(0.000)**	**(0.000)**
Mother's Educ>9 Mean	1.000	1.003	1.012
Mother's Educ≤9 Mean	1.000	1.000	1.002
p-value	**(0.000)**	**(0.000)**	**(0.000)**
Grandmother's Educ>3 Mean	1.000	1.002	1.005
Grandmother's Educ ≤3 Mean	1.000	0.999	1.002
p-value	**(0.000)**	**(0.000)**	**(0.000)**

1. About 40% of children's grandmothers have more than 3 years of formal education.
2. *p*-values are presented in the parentheses.

Table J.8: The Effects of Interaction Measures on Fine Motor Skill Tasks' Learning Components by Fine Motor Skill Ability

Mean	Teaching Ability	Interaction Quality Home Visitor and Caregiver	Interaction Quality Home Visitor and Child
Slow Mean	1.000	0.987	0.973
Fast Mean	1.000	1.007	1.016
p-value	**(0.000)**	**(0.000)**	**(0.000)**
Normal Mean	1.000	0.997	0.997
Fast Mean	1.000	1.007	1.016
p-value	**(0.000)**	**(0.000)**	**(0.000)**
Slow Mean	1.000	0.987	0.973
Normal Mean	1.000	0.997	0.997
p-value	**(0.000)**	**(0.000)**	**(0.000)**

1. Ability Definition:

Fast group: child passes the first tasks for more than 80% difficulty levels, and the average passing rate is more than 80% of all the tasks. Normal group: Child passes the first tasks less than 80% of difficulty levels, and the average passing rate is between 50% and 80%. Slow group: the average passing rate is less than 50%.

2. *p*-values are presented in the parentheses.

Table J.9: The Effects of Interaction Measures on Fine Motor Skill Tasks' Learning Components by Fine Motor Skill Ability and Age of Enrollment

Ability		Teaching	Interaction Quality	Interaction Quality
	Age enrollment 20–25 Months Mean	1.000	1.006	1.016
	Age enrollment 9–15 Months Mean	1.000	1.005	1.015
	p-value	**(0.021)**	(0.192)	(0.731)
	Age enrollment 15–20 Months Mean	1.000	1.007	1.016
Fast	Age enrollment 9–15 Months Mean	1.000	1.005	1.015
	p-value	**(0.009)**	**(0.000)**	(0.331)
	Age enrollment 20–25 Months Mean	1.000	1.006	1.016
	Age enrollment 15–20 Months Mean	1.000	1.007	1.016
	p-value	(0.825)	**(0.000)**	(0.591)
	Age enrollment 20–25 Months Mean	1.000	0.998	1.004
	Age enrollment 9–15 Months Mean	1.000	0.998	0.996
	p-value	(0.923)	(0.190)	**(0.000)**
	Age enrollment 15–20 Months Mean	1.000	0.997	0.993
Normal	Age enrollment 9–15 Months Mean	1.000	0.998	0.996
	p-value	**(0.000)**	**(0.071)**	**(0.015)**
	Age enrollment 20–25 Months Mean	1.000	0.998	1.004
	Age enrollment 15–20 Months Mean	1.000	0.997	0.993
	p-value	**(0.000)**	**(0.004)**	**(0.000)**
	Age enrollment 20–25 Months Mean	1.000	0.986	0.976
	Age enrollment 9–15 Months Mean	1.000	0.987	0.967
	p-value	(0.106)	(0.635)	**(0.001)**
	Age enrollment 15–20 Months Mean	1.000	0.987	0.972
Slow	Age enrollment 9–15 Months Mean	1.000	0.987	0.967
	p-value	**(0.007)**	(0.665)	(0.126)
	Age enrollment 20–25 Months Mean	1.000	0.986	0.976
	Age enrollment 15–20 Months Mean	1.000	0.987	0.972
	p-value	**(0.000)**	(0.379)	(0.103)

1. *p*-values are presented in the parentheses.

2. Ability Definition:

Fast group: child passes the first tasks for more than 80% difficulty levels, and the average passing rate is more than 80% of all the tasks. Normal group: Child passes the first tasks less than 80% of difficulty levels, and the average passing rate is between 50% and 80%. Slow group: the average passing rate is less than 50%.

References

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Heckman, J. J. and E. J. Vytlacil (2007a). 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.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs, and to forecast their effects in new environments. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 71, pp. 4875–5143. Amsterdam: Elsevier B. V.

Matzkin, R. L. (1993, July). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics 58*(1–2), 137–168.

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

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

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