# A Study of the Microdynamics of Early Childhood Learning*

James Heckman and Jin Zhou

January 9, 2025

*This research was supported in part by NIH grant NICHD R37HD065072, RGC grant 9048312, and an anonymous donor. We thank our partners on the China REACH project: China Development Research Foundation, Beijing Normal University, the 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 and Hong Kong Research Grants Council. 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), at Rice University (November, 2021), at the Cowles conference on Labor and Public economics (June, 2023), at City University Hong Kong (June, 2024), at Society of Economic Dynamics in Barcelona (June, 2024), and the Asian Meetings of the Econometric Society in Hangzhou (June, 2024). We have benefited from comments received in four rounds of revision over four years under the the aegis the editor, John List, and four referees, as well as those of our discussant Lance Lochner at a September 2019 Becker Friedman Institute conference in Chicago. Flavio Cunha and Pat Kyllonen gave us valuable commentary. We have also greatly benefited from the research and commentary of Haihan Tian and Zijian Zhang who are coauthors on related papers. Fuyao Wang and Alejandra Campos also contributed highly competent and insightful research assistance and commentary.

**Abstract**

This paper investigates the weekly evolution of child skills as measured by unique data from a widely-emulated early childhood home-visiting program developed in Jamaica, adapted to rural China, and applied in different versions worldwide. The design of the study avoids problems of endogeneity of inputs and lack of truly comparable measures of skills across children that plague previous econometric studies of child development. Skills that are nominally classified as the same, in fact, do not appear to share a common unit scale across levels. They are produced by skill-specific, lifecycle-stage-specific technologies. We formulate and estimate a new dynamic stochastic skill production model for multiple skills that is consistent with the evidence. We quantify the dynamics of early life learning. The model explains the "fadeout" of measures of learning sometimes attributed to forgetting or depreciation of skills. We investigate the role of ability in learning. We find important differences in learning patterns between boys and girls.

**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
Department of Economics
and Finance
City University of Hong Kong
9-223 Lau Ming Wai Building
Kowloon
Hong Kong
Email: jin.zhou@cityu.edu.hk

# 1 Introduction

This paper uses weekly measurements of skills on children in a prototypical home visiting program, implemented at scale in China, to investigate the mechanisms producing the growth of multiple skills at early ages. The design of our sample allows us to bypass concerns about input endogeneity and the incomparability of measures of skill across people and over ages for the same person.[1] Access to detailed weekly data enables us to determine at what lifecycle stages learning occurs, at what rate, and how family environments affect it.

We develop and estimate a micro-dynamic model of learning that characterizes the evolution of skills during early childhood. It is a model of reinforcement learning that differs substantially from standard models of skill formation used in the current literature. We measure the impact of information provided to parents on boosting children's skills. Different levels of nominally the same skill are characterized by different production functions.

Versions of the *technology of skill formation* (Cunha and Heckman, 2007; Cunha et al., 2010) are currently widely used to characterize the growth of child skills $\boldsymbol{K(a)}$ at age (stage) $\boldsymbol{a}$. These technologies are functions of a vector of investments $\boldsymbol{I(a)}$ (parenting, other interactions with the child by childcare workers, etc.) and environments $\boldsymbol{G(a)}$ (including neighborhoods, peer effects, parental education, and public goods, such as schooling, as in Agostinelli et al., 2022):

[1] See, e.g., Cunha et al. (2021) on the issue of the arbitrariness in scales of test scores. See also Cawley et al. (1999) and Bond and Lang (2013).

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

The technology is age-specific, inputs are normalized so that output increases in each argument. It is usually assumed to be twice differentiable.

Properties of this technology are exposited in Heckman and Mosso (2014). A recurrent finding of the literature is that enhancements in parenting are associated with improvements in child outcomes (García and Heckman, 2023). This paper studies the impact of a parenting intervention on the growth of child skills. The intervention promotes parenting, but with different effects for children with different types of parents and home environments. We study the dynamic impacts of the program as mediated by these factors.

In addition, we also address the question of how to measure skills and their growth when scales of skills are arbitrary, and hence, comparisons over time and across persons are problematic. In the literature, test scores based on assessments of cognitive, socioemotional, and other skills are widely used.[2] It has long been noted that such measures have intrinsically arbitrary scales (e.g., Uzgiris and Hunt, 1975; Cunha and Heckman, 2008; Cunha et al., 2010, 2021). Ordinal production functions that compare ranks across people do not suffer from this problem but, at the same time, do not provide interpretable measures of levels of attained skill.[3] Freyberger (2022) shows the dramatic consequences of different scalings of skill measures for

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

[3]Cunha et al. (2010, 2021); Agostinelli and Wiswall (2021); Bond and Lang (2013); Freyberger (2022).

estimates of technology.

This paper presents empirical evidence on the learning process. We study the impacts of home visits that not only teach the skill specific tasks, but also inform parents about effective parenting strategies and the impact of home visitor quality. We examine how home environments mediate the impacts of these investments. We lack data on the specific nature of the induced caregiver-child interactions resulting from home visiting with caregivers.

We develop and estimate a latent Markov model of skill formation that explains the growth of measured skills and explains why the growth is not necessarily monotonic with respect to exposure to the program. We formalize intuitive models of child development used in psychology.[4] We investigate the growth of skills at far more granular levels than previous analyses in economics or psychology.

We address the problem of the arbitrariness of test scores by using scales of skills constructed to be comparable *within* well-defined levels of skills but not necessarily across levels of skills. We do not impose a common scale of skills across levels of nominally the same skill as is traditionally done in the literature.[5]

We report the following findings. (1) Our estimated technology is skill and lifecycle-stage-specific. The estimated technology differs greatly across levels of nominally the same skill. (2) Investment *in caregivers* by home visitors promotes the growth of skills of children; (3) The impact of this investment is mediated by caregiver and home visitor traits. Grandparents and parents with less education

[4]See, e.g., Bronfenbrenner (2005) and Thelen (2005). See also Bailey et al. (2020).

[5]See, e.g., Todd and Wolpin (2007); Cunha and Heckman (2008); Cunha et al. (2010); Attanasio et al. (2020).

apparently provide less stimulation in response to the intervention than more educated caregivers;[6,7] (4) Stocks of skills cross-fertilize the growth of other skills but not symmetrically; (5) Investment in different skills exhibits cross-productivity for some skills but not others; and (6) There are gender differences in the dynamics of learning.

Because we lack details on the exact nature of parental responses to home visits, we do not measure all the channels through which home visits operate. Nonetheless, we can assess the effects of different home environments on the home visit received.

The paper unfolds in the following way. Section 2 describes the background of the program we analyze and its curriculum.[8] Section 3 presents evidence of learning patterns induced by it. Section 4 presents a latent Markov learning model for skills that is concordant with the evidence. Section 5 presents estimates and interpretations. Section 6 concludes.

## 2 China REACH

The inspiration for the program we analyze is the Jamaican Home Visiting Intervention (Grantham-McGregor and Smith, 2016), a randomized home visiting parenting intervention given to a sample of 129 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

[6] In our sample, grandparent caregivers have on average three years of education while parents and home visitors have roughly ten years of schooling.

[7] Heckman et al. (2024) present a more nuanced nonparametric analysis of this point.

[8] Zhou et al. (2024) describe it in much greater detail.

Bangladesh, China, Colombia, India, and Peru (see, e.g., Grantham-McGregor and Smith, 2016).

The program we analyze, *China REACH*, extends and applies the Jamaican protocols. Implemented in 2015 by a large-scale randomized control trial, it enrolled 1,500 subjects (age 6 months-42 months) in 111 villages in Huachi county, Gansu province, one of the poorest areas of China. Unlike the original program, this intervention is not focused on stunted children. Severely impaired children do not participate.

China REACH is a paired-matched RCT that minimizes the mean square errors of estimates (Bai et al., 2021; Bai, 2022). A non-bipartite Mahalanobis matching method was used to pair villages and randomly select one village within a pair into the treatment group and the other village into the control group.[9] More details of the design of the experiment and balance tests for treatment and control groups can be found in Zhou et al. (2024).

The intervention cultivates multi-dimensional skill development through home-visiting. Trained home visitors who are roughly at the same level of education of the mothers of the children studied visit each treated household weekly and provide one hour of caregiving guidance.

Zhou et al. (2024) evaluate the treatment effects of the intervention using a different inventory of outcome measures than the ones used here. Only two measurements are collected at midline and endline of the intervention in contrast with the weekly measurements analyzed here. They find that the intervention significantly improves

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

skill development (e.g., language and cognitive, fine motor, and social-emotional skills). To interpret treatment effects, they use item responses on measures of skill to estimate individual latent skills. They decompose treatment effects and find that enhancement of latent skills explains most of the estimated conventional treatment effects. Zhou et al. (2023) show that the skill profiles for the growth of skills are similar to those of the original Jamaica Home Visiting program over ages where comparable data exist, suggesting the applicability of our analysis to the original program. Heckman et al. (2024) present evidence on dynamic complementarity. The focus in this paper is on the growth of skills in the treatment group, and not on treatment effects per se.

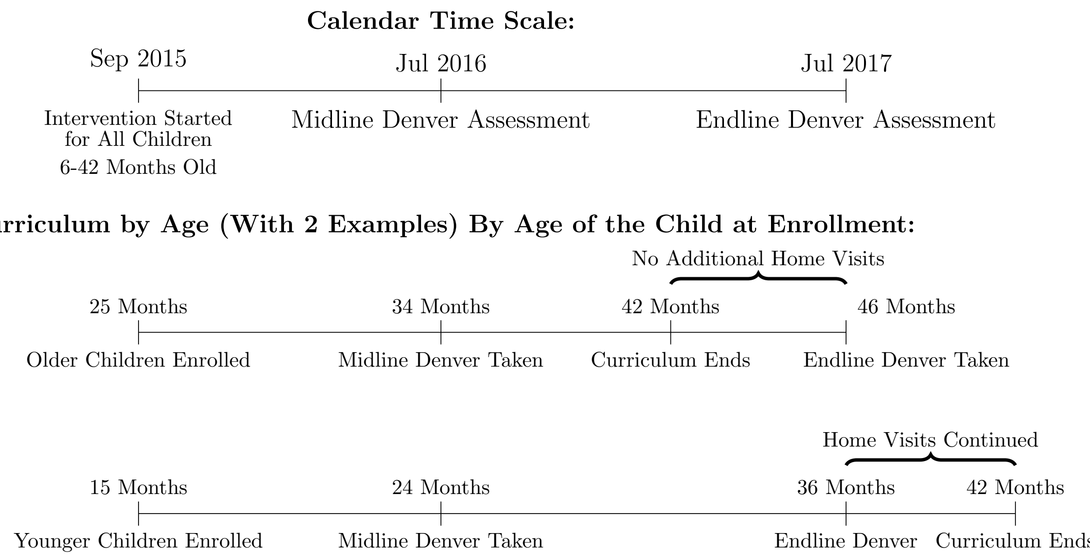
## 2.1 Enrollment Protocol

The program enrolls all children age 6-42 months as of September 2015. Figure 1 gives the enrollment time frame, and the Denver assessment timing analyzed in Zhou et al. (2024). It shows that different cohorts defined by age get different exposures to the program. All caregivers of children of the same age in the program get the same lesson at the same age. Children are evaluated weekly on their knowledge. The lessons given are exogenously determined. Visitors are chosen from the target villages and are essentially homogenous across villages and of the same level of education as the mothers visited. They are essentially randomly assigned.[10] However, the

[10] According to the information collected by the CDRF field team, 50 villages out of 55 villages have an average of 9 years of education for home visitors, which is about 90% of all treated villages. For two villages, the average years of education for home visitors are about six years; for two villages, it is about 12 years; and for one village, it is about 14 years. For Pearson $\chi^2$ statistic ($\chi^2(54) = 9.50$), we cannot reject the null hypothesis that all the villages have the same distributions of years of education for the home visitors. When we remove the anomalous villages, we get essentially the

implementation of the lessons received depends on caregivers and home environments. Older children at entry do not get the training that earlier entrants receive. There is no attrition from the program (except by death).

Figure 1: China REACH Calendar Time Scales




The Denver assessments taken at midline and endline measure child development for both treatment and control children. They are not analyzed in this paper.[11] We focus on the growth of skills in the treatment group.

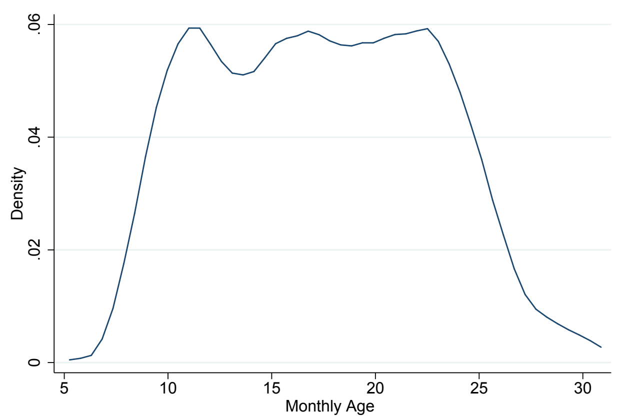
Figure 2 plots the distribution of the age of entry into the program in September 2015 of different age cohorts. The cohorts are more or less randomly distributed between 10-25 months old. Table A.1 in Appendix A documents the balance in

same results. The Pearson statistic is $\chi^2(49) = 12.72$ after removing the anomalous villages. We cannot reject the null hypothesis that all the villages have the same years of education. Estimates are essentially the same with and without inclusion of these villages.

[11]They are analyzed in Zhou et al. (2024).

backgrounds across different enrollment cohorts. Few children older than 25 months are enrolled.

Figure 2: The Distribution of Monthly Age when Enrolled into the Program




## 2.2 Program Protocols

The program teaches and encourages the caregiver to interact with the child through playing games, making toys, singing, reading, and storytelling to stimulate the child's cognitive, language, motor, and socioemotional skill development. The home visit to the caregiver is the intervention studied. We lack data on the precise way caregivers act on the information they receive. Using a rich set of observed caregiver characteristics, we estimate how caregivers with different educational attainment and background mediate the impact of home visits on child development.

Four different skill tasks (gross motor, fine motor, language, and cognitive) are taught each week. Skills taught are ordered by difficulty levels following profiles developed by Palmer (1971) and Uzgiris and Hunt (1975), henceforth UHP.[12] These scales are widely used in the literature on child development. They are the ones analyzed in this paper. The intervention instructs caregivers at the weekly level on activities to promote the skills that appear in these scales.[13] The caregiver is the vessel, and as we shall see, different caregivers have different effectiveness in promoting child development.

Central to our identification strategy is the use of scales that describe valid levels of knowledge with knowledge content that is the same *within* each level.[14] Child skills are assessed weekly. There are monthly assessments of the quality of home visits recorded by supervisors, and data on the quality of home environments are also collected.

There are 13 difficulty levels for cognitive skills. Table 1 gives the tasks for cognitive skills taught at specific levels, and Figure 3 presents the timing of the lessons taught by age. The tasks start with simply understanding a picture by verbal acknowledgment to using receptive (heard) language to identify pictures.

Although task content progresses by levels, it is designed to be essentially identical *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 activity of looking at pictures or objects and vocalizing. Appendix C gives comparable information for the other skills which follow the same pattern.

[12]More details about the curriculum are provided in Appendix B.

[13]Some of these scales also appear in the Denver test.

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

Table 1: Difficulty Level List for Cognitive Skill Tasks

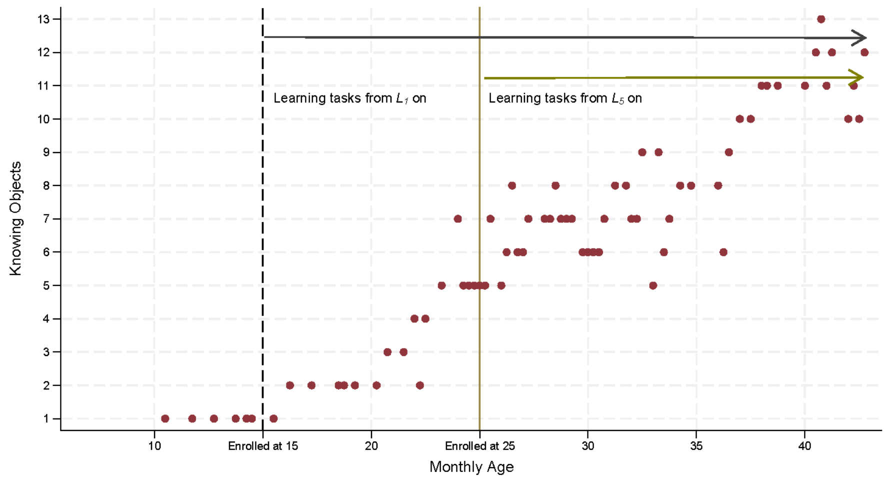
| | |
|---|---|
| Level 1 | Look at the pictures and vocalize |
| Level 2 | Name the objects and ask the baby to point to the pictures accordingly |
| Level 3 | The child can name the objects in one picture, and point to the named picture |
| Level 4 | The child can name the objects in two or more pictures, 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 six or more pictures |
| Level 7 | The child can talk about the pictures, answer questions, understand, or name the verbs (eat, play, etc.) |
| Level 8 | The child can follow the storyline, name actions, and answer questions |
| 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 graph, 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 |

Source: Scales are from Wachs et al. (1971).

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 |

Figure 3: The Timing of Teaching Cognitive Skills (Understand Objects) Tasks across Difficulty Levels and Two Possible Enrollment Patterns




Note: Level 1: look at the pictures and vocalize; Level 13: the child can name things in the pictures and talk about the the functions of the objects

The fact that the skills taught and assessed *within levels* are essentially identical is crucial to our approach.

# 3 Evidence on Learning

To understand the structure of the data analyzed, it is helpful to introduce some notation. Let $\mathcal{S}$ be the set of skills taught. Let $\ell(s, a)$ be the level of skill $s$ taught at age $a$. Within levels, skills are identical. At the outset of each weekly visit, the home visitor records a binary measure of whether the child can master the task previously taught (i.e., whether the child understands the task previously taught). For skill $s$, at difficulty level $\ell$, and weekly age $a$, the task item is uniquely determined in the curriculum. We use $D(s, \ell, a)$ to denote whether or not a child knows the task associated with latent skill $s$ and level $\ell$ at age $a$, $K(s, \ell, a)$, which we characterize 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 $D(s, \ell, a)$ is the data we observe, recording knowledge of skill $s$ at level $\ell$ at a given level at age $a$. $\bar{K}(s, \ell)$ is the minimum level latent skill required to accomplish the task at difficulty level $\ell$. It is the same for all tasks within level $\ell$ for each $s$ by construction.

This characterization is similar to that used in the classical IRT model (Lord and Novick, 1968) and models of discrete choice (Thurstone, 1927; McFadden, 1981). Define $\underline{a}(s, \ell)$ as the first age at which skill $s$ is taught at level $\ell$, and let $\bar{a}(s, \ell)$ be the last age at which it is taught at level $\ell$. For level $\ell$ of skill $s$, indicators of knowledge in a spell are elements of: $\left\{D(s, \ell, a)\right\}_{\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}$. For example, for cognitive skill level one, $\underline{a}(s, \ell)$ is age 10 months and 2 weeks and $\bar{a}(s, \ell)$ is age 15 months and 2 weeks. Seven tasks at level one were taught during this age range. Therefore, in our data, we observe seven indicators to record whether the child had knowledge (or not) of skill $s$ at age $a$.

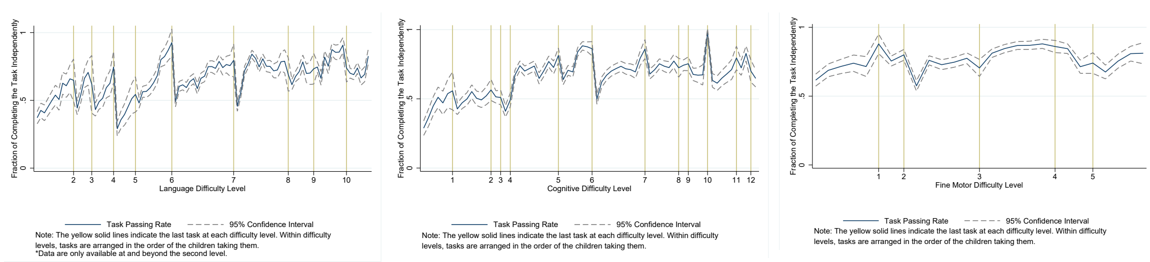
The sample passing rate on the test for skill $s$ at level $\ell$ at age $a$ is the mean of $D(s, \ell, a)$ for children tested on the age $a$ item for the skill $s$. It is the mean passing rate for the item. $Pr(D(s, \ell, \bar{a}(s, \ell)))$ is a measure of final skill $s$ level attainment in level $\ell$.

## 3.1 Patterns of Learning

Figure 4 plots the growth of knowledge in language, cognitive, and fine motor skills.[15] Average (across people) passing rates by age within each difficulty level for language and cognitive tasks increase with age, a pattern consistent with learning. When individuals transition to higher difficulty levels, initial age-specific passing rates decline. This is consistent with the notion that new skills are taught at each level.[16] After initial declines, age-specific passing rates within levels increase as learning ensues. The dynamic model presented in Section 4 below captures this phenomenon. At most levels of fine motor skills, there is—at best—modest learning. Access to detailed weekly data enables us to determine at what stages learning occurs, at what rate, and how family environments and caregiver-home visitor interactions affect it.

Figure 4: Average Task Passing Rates by Order and Level

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




Source: See primary data and the plots in Zhou, Heckman, Wang, and Liu (2023).

[15] We also measure gross motor skills, but they are not affected by the intervention (Zhou et al., 2024), so we do not systematically analyze them here.

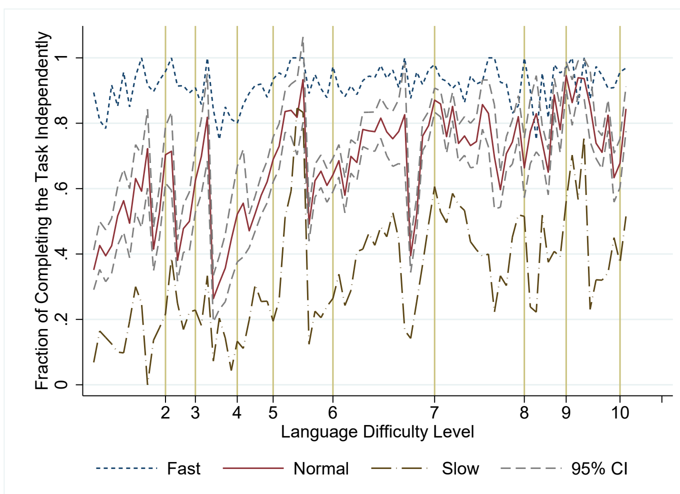
[16] Alternatively, this might arise if the difficulty levels of assessments for the same skill increase across levels. There is nothing in program design that increases difficulty levels of the assessments in this fasion.

Table 3: Ability Categories (Measured over All Levels)

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

Figure 5 disaggregates Figure 4 by ability as defined in Table 3, as the speed of learning *across all levels*. There is high persistence of this measure of ability across difficulty levels for the same skill. See Appendix D for a detailed discussion of this measure. Low-ability children learn more slowly.

Figure 5: By Ability Group: Average Language Task Passing Rates




The sawtooth patterns arise from the transitions across levels for language skills. The pattern for normal and low-ability children is consistent with the notion that a new type of skill is being learned across transitions. High-ability group children, on average, have the highest passing rate, a phenomenon that persists across levels and is found for other skills (See Appendix E, Figures E.1-E.4).

# 4 Mechanisms Generating Child Learning

To motivate our approach to estimating the weekly dynamics of skill formation, we consider a simple version of the model for one level of one skill before presenting our general model. The more general model is the simple model applied to each skill at each level, with parameters that may vary across levels and skills. We then consider a model of joint skill formation.

We use the notation previously introduced in Section 3, but suppress the $s$ and $\ell$ because we initially only consider one skill at one level. The program fosters skill at ages $a \in [0, \ldots, \bar{A}]$. Lessons are the same for all participants at age $a$. We define $K(a)$ as the level of "skill" achieved 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 *knowledge* of the skill at level $\bar{K}$ at age $a$ if $K(a) \geq \bar{K}$. $D(a) = 1$ if a person at age $a$ masters the skill, so $D(a) = \mathbf{1}\big(K(a) \geq \bar{K}\big)$. Skill level is measured at each age.

We assume i.i.d. idiosyncratic shocks in growth rates ($\varepsilon(a)$) on a log scale. A multiplicative version of the model turns out to fit the data on skill growth very well.[17] Skill acquisition is characterized as a random walk:

$$\ln K(a) - \ln K(a-1) \doteq \delta(a)\eta + V(\boldsymbol{Q}(a)) + \varepsilon(a). \tag{3}$$

$\eta$ is ability to learn the skill. It is assumed to be individual specific and positive ($\eta > 0$), and $\delta(a)$ is the "lesson" at age $a$ for all children enrolled in $a$. $V(\boldsymbol{Q}(a))$ captures variables $\boldsymbol{Q}(a)$, such as family background and investments received at home, as well as autogenic effects that affect the evolution of skills.[18] $V(\boldsymbol{Q}(a))$ also operates independently of the level of $\ln K(a-1)$. We assume a common scale of skill *within* each designated skill level. Skills are assumed to be additive in the metric that quantifies $\ln K$.

Accounting for initial conditions, we can write Equation (3) as:

$$\ln K(a) \doteq \underbrace{\eta \sum_{j=1}^{a} \delta(j) + \sum_{j=1}^{a} V(\boldsymbol{Q}(j)) + \sum_{j=1}^{a} \varepsilon(j)}_{L(a)} + \ln K(0) \tag{4}$$

where $\varepsilon(j)$ is i.i.d. across all $j$ with $E(\varepsilon(j)) = 0$. Random walk growth in skills was introduced in Rutherford (1955).

This economic model extends models in psychometrics, in particular the Item Response Theory (IRT) model (Lord and Novick, 1968), that measure skills at a

[17]Appendix G compares the empirical performance of multiplicative and additive models. The additive model uses $K$ in place of $\ln K$ in Equation (4). In many aspects, the qualitative results from each are very similar, but quantitative results are somewhat better for the multiplicative model as characterized by goodness of fit and model specification tests.

[18]By autogenic effect, we mean growth not directly attributable to the program, e.g., imitation, peer effects, etc. Recall that children enter at different ages and may have different levels of preprogram environmental exposures.

point in time. An essential feature of the IRT model is captured by the threshold crossing feature (2).[19] Because of the random walk component in (4), we generalize the stochastic properties of the IRT model which assumes independence in outcomes conditional on a scalar unobservable, usually interpreted as "ability." In our setup, ability grows across learning occasions unlike in the IRT model.

In this notation, *self productivity* is:

$$\frac{\partial K(a+1)}{\partial K(a)} = exp\,\{\underbrace{\delta(a)\eta + V(\boldsymbol{Q}(a)) + \varepsilon(a)}_{J(a)}\}.$$

*Investment productivity* is:

$$\frac{\partial K(a+1)}{\partial \delta(a)} = K(a)\eta\; exp\;(J(a)).$$

*Static complementarity* is:

$$\frac{\partial^2 K(a+1)}{\partial K(a)\partial \delta(a)} = \eta\; exp\;(J(a)).$$

*Dynamic Complementarity* is:

$$\frac{\partial^2 K(a+j+1)}{\partial \delta(a+j)\delta(a)} = K(0)\eta^2\; exp\;(L(a+j+1)).$$

If $\eta > 0$, both static complementarity and investment productivity are positive.

[19] The Bayesian Knowledge Tracing (BKT) model is captured by the dynamics of the model of Equation 3. Unlike the BKT model, knowledge $K(a)$ in our model is affected by education and investment, which is captured by $\delta(a)$, so that we depart from its mechanical growth trajectory feature to account for investment that affects learning. Deonovic et al. (2018) compare the IRT and BKT models and criticize them for not including investment as a determinant of learning.

Adding stochastic shocks to learning growth allows for either growth or shrinkage around deterministic growth paths. Shrinkage could be due to forgetting or distraction on the test. The decline in latent knowledge is sometimes called "fadeout." The literature on fadeout of test scores (see, e.g., Bailey et al., 2020) assumes deterministic growth profiles, whereas we allow for stochastic growth and fadeout of measured skills within a lifetime.

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

$$\Pr(D(a) = 1 \mid \eta, \Delta(a), \Lambda(a), K(0)) = \int_{\ln \bar{K} - \eta\Delta(a) - \Lambda(a) - \ln K(0)}^{\infty} dF_a(U(a)), \tag{5}$$

where $F_a$ is the cdf of $U(a)$.

## The General Model for Scalar Skills

The general model has the same structure as the simple model applied to skills at each level where $\boldsymbol{\mathcal{S}}$ is the set of skills taught, $\ell(s, a)$ is the level of skill $s$ taught at age $a$, and there are $L_s$ levels of difficulty for each skill $s$.

Shocks at level $\ell$ for age $a$—$\varepsilon_\ell(s, a)$—are assumed to be independent across $a$. Their distributions may vary with $\ell$ and $s$. When estimating the model, we assume that they are i.i.d. within $\ell$ for each skill $s$, and across $s$, but not necessarily across

$\ell$, $\eta(s)$ may vary by age $a$[20] and $\delta(a)$ captures the content of the home visiting curriculum. Thresholds (passing standards) $\bar{K}(s,\ell)$ may also change across levels, as may $V_\ell(\boldsymbol{Q}(a))$.

By allowing for level-specific shocks, we account for the possibility that different difficulty levels within an assessment may have different variances. This is, indeed, what we find in our estimates. We can explain the decline of measured skills within a lifetime by allowing for shocks $\varepsilon(s,\ell)$ within and across levels and differences in difficulty across levels.

## 4.1 Testing for A Common Scale of Skills Across Skill Levels within the Model

This paper develops and applies a model-based test of a common scale of skills across levels. By this, we mean that the scale of nominally the same skill is the same across different difficulty levels, a common assumption in human capital models since Ben-Porath (1967).[21]

Under the common scale assumption across levels, latent index $\ln K(s,\ell,a)$ cumulates so measures of knowledge growth are well-defined, at least at the level of latent skills. This requires, among other things, that in the absence of depreciation

[20] In our estimates, $\eta$ includes the interaction measures and a measure of grandmother's education when she is the caregiver. Therefore, $\eta$ changes as lessons change.

[21] Cunha et al. (2010) impose a common scale assumption. In our dynamic learning model, we do not need to impose a common scale assumption. Our weekly data are richer, which makes it feasible to test the common scale assumption. Cunha et al. (2010) also impose linearity on aggregate scores. Our learning model estimates using item level data. This is a more nuanced approach for studying the learning process.

(or appreciation) associated with transitions across levels,

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

This is a property of latent variables *at the junction points of levels.* Measurement of these skills is an entirely separate matter. We test for a common scale across levels, maintaining the assumption of a common scale *within levels*. Our proof of model identification in Section 4 Appendix H makes this point precise. The assumed lack of depreciation (or appreciation) is a property that holds only at junction points across levels, and not at all ages within levels, which would impose a lack of growth on the model.

If scales change across levels, but human capital scales are somehow connected, we write:

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

where $\Gamma_\ell$ is a general function. If there is total depreciation of skills in that transition from $\ell - 1$ to $\ell$, $\Gamma_\ell$ is the zero function. The property of a common scale across the junction between $\ell$ and $\ell - 1$ sets $\Gamma_\ell = I$ (no depreciation **or appreciation**)–the identity function. Depreciation of the same skill across junction point $\ell$ is $\Gamma_\ell = 1 - \sigma_\ell(s)$, where $\sigma_\ell(s)$ is depreciation at level $\ell$ for skill $s$. $\sigma_\ell(s)$ can be negative so there can be appreciation.[22] This paper only considers affine transformations for $\Gamma_\ell(\cdot)$:

[22]This is a one-shot markdown or markup of skill across levels.

$$\Gamma_\ell(K(s,\ell,\underline{a}(s,\ell))) = \gamma_{0,\ell} + \gamma_{1,\ell}(K(s,\ell,\bar{a}(s,\ell))). \tag{6}$$

We use an affine transformation as a first-order linear approximation of a general function. Setting $\gamma_{0,\ell} = 0$ and $\gamma_{1,\ell} = 1$ captures the notion of a common scale in the absence of depreciation. With depreciation, $\gamma_{1,\ell} = 1 - \sigma_\ell(s)$ (i.e., $\sigma_\ell(s) > 0$), a one-shot change in skill level after crossing the boundary. Similarly, with appreciation, $\gamma_{1,\ell} = 1 - \sigma_\ell(s)$ (i.e., $\sigma_\ell(s) < 0$).

Notice that we are testing how latent skills are connected across levels of nominally the same skill, but we do not impose linearity on the skill formation process. The common scale assumption would be violated if new skills emerge at each level, or if a new transformation of skills would be relevant.

## 4.2 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 H. We use means and covariances because we assume normal shocks in estimation. In the appendix, we show that we can nonparametrically identify the joint distributions of unobserved variables up to normalizations.

Define the latent index $\ln K(1,a)$ for skill at level 1 at age $a$. This corresponds to $\ln K(s,1,a)$ for a particular skill $s$, which is kept implicit. We simplify Equation (4) to read:

$$\ln K(1,a) = \underbrace{\eta \sum_{j=1}^{a} \delta_1(j)}_{\text{learning}} + \underbrace{V_1(a)}_{\substack{\text{autogenic}\\ \text{growth}}} + \underbrace{U_1(a)}_{\text{shocks}} + \ln K(0), \tag{7}$$

where $\ln K(1,a)$ is the latent index (skill) at difficulty level 1 at weekly age $a$, and $K(0)$ is the initial condition. We assume $\ln K(0) = \mu_0(\boldsymbol{Z}) + \Upsilon$, where $\boldsymbol{Z}$ 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 $\delta_1(a)\eta(\boldsymbol{X}) = \bar{\beta}_1(\boldsymbol{X}) + \omega$, where the $\boldsymbol{X}$ are covariates, including various interactions, background variables, and gender indicators. We assume that $\boldsymbol{X} \perp\!\!\!\perp [\omega, \varepsilon_1(j)]$ for all $j$. $\omega$ 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, if we do so, we have to take a position on the dependence across $\omega_j$.[23] 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 $\sum_{j=1}^{a} V_1(\boldsymbol{Q}(j))$.

Equation (7) can be rewritten in the notation for the general case allowing for heterogeneity in $\ln K(0)$:

$$\ln K(1,a) = \mu_1 + \mu_0(\boldsymbol{Z}) + V_1(a) + \bar{\beta}_1(\boldsymbol{X})a + \underbrace{\left\{ a\omega + \sum_{j=1}^{a} \varepsilon_1(j) + \Upsilon \right\}}_{\Psi_1(a)} \tag{8}$$

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

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

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

where $\mu_1^* = \mu_1 - \bar{K}(1)$ and $\mu_1$ 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 $\mu_1^*$, $\mu_0(\boldsymbol{Z})$, $\bar{\beta}_1(\boldsymbol{X})$, $V_1(a)$ up to scale. Since we can identify the ratio of $\frac{\sigma(1,j)}{\sigma(1,j')}$, $\sigma(1,a)$, $\sigma(1,a')$ are identified up to a normalization (e.g., $a, a' \neq j$) (see Heckman, 1981 and Heckman and Vytlacil, 2007). We discuss the time varying components of $\boldsymbol{X}$ in our data in the next section when we discuss empirical estimates.

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 scale after normalizing $\sigma^2(1,j) = 1$. On the right-hand sides, there are three unknown terms $\sigma^2_\omega$, $\sigma^2_{\varepsilon(1)}$, and $\sigma^2_\Upsilon$. When $a \geq 3$ (i.e., three different tasks at level one), we can identify all three terms: $\sigma^2_\omega$, $\sigma^2_{\varepsilon(1)}$, and $\sigma^2_\Upsilon$ with sufficient variation in $a$ and $j$.

Adopting a similar notation for levels $\ell > 1$, if we assume a common scale of skills across level 1 and 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:

$$\begin{aligned}
\ln K(2,a) &= \mu_2 + V_2(a) + \bar{\beta}_2(\boldsymbol{X})(a - \bar{a}(1)) + \sum_{j=\underline{a}(2)}^{a} \varepsilon_2(j) + \ln K(1, \bar{a}(1)) \\
&= \mu_1 + \mu_2 + \mu_0(\boldsymbol{Z}) + V_1(\bar{a}(1)) + V_2(a) + \bar{\beta}_2(\boldsymbol{X})(a - \bar{a}(1)) + \bar{\beta}_1(\boldsymbol{X})\bar{a}(1) \\
&+ \underbrace{\left\{ \sum_{j=\underline{a}(2)}^{a} \varepsilon_2(j) + (a - \bar{a}(1))\omega + \sum_{j=1}^{\bar{a}(1)} \varepsilon_1(j) + \bar{a}(1)\omega + \Upsilon \right\}}_{\Psi_2(a)}. \qquad (9)
\end{aligned}$$

Given the initial normalization at level 1 (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 $\bar{\beta}_2(\boldsymbol{X})$ up to scale $\sigma(2,a)$, where

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

Since we have already established identification of $\sigma_\omega^2$, $\sigma_{\varepsilon(1)}^2$, and $\sigma_\Upsilon^2$, the only term not identified in $\mathrm{Var}\Psi_2(a)$ is $\sigma_{\varepsilon(2)}^2$. We now discuss how to identify this term. Consider the covariance term $\mathrm{Cov}\left(\frac{\Psi_2(a)}{\sigma(2,a)}, \frac{\Psi_2(a')}{\sigma(2,a')}\right)$

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

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 $\sigma^2_{\varepsilon(2)}$.[24] 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 $\bar{\beta}_2(\boldsymbol{X})$ up to $\sigma(2,a)$.

From the previous discussion for all $\ell \geq 2$, we can identify the variance $\sigma(\ell,a)$ without imposing additional normalization at levels $\ell$ ($\ell \geq 2$). The only normalization we need is on the scale of variance term $\sigma(1,j)=1$ at level 1.[25]

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 H.5. We do not develop this point further because we adopt parametric models in forming our estimates. The conditions just developed extend in a straightforward way to higher levels, $\ell > 2$, and to the multivariate model discussed below. All higher-level parameters are identified up to the initial normalization at level 1.

[24]We take positive roots in solving the implicit quadratic equation.

[25]We can impose any maintained value of $\sigma_\ell^{(s)}$.

### 4.2.1 Testing the Common Scale Assumption

Under an assumption of a common scale of skills characterized by Equation (6) with $\gamma_{0,\ell} = 0$ and $\gamma_{1,\ell} = 1$, we obtain tight restrictions on the coefficients across levels. Relaxing this assumption adds two new parameters $(\gamma_{0,2}, \gamma_{1,2})$ to Equation (9):

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

Notice that the common scale assumption in the form we use it imposes a proportionality restriction across functions common to $\ln K(2,a)$ and $\ln K(1,a)$. Going across levels,

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

From the previous analysis, the term in braces is identified up to the previously stated normalization at the first level. Thus $\gamma_{1,2}$ is identified, and we can test if $\gamma_{1,2} = 1$. We can use this logic to identify depreciation operating across junction points if we maintain a common scale assumption.

Testing $\gamma_{0,2} = 0$ requires stronger assumptions. We need model intercepts to be invariant across levels, which is difficult to maintain given that $\bar{K}(2)$ is absorbed in any estimated intercept. We expect that the difficulty levels are increasing in $\ell$. As before, we can estimate $\ln \bar{K}(2)$ up to scale net of intercepts, and we can identify the scale. We impose $\gamma_{0,2} = 0$ without loss of generality because it is absorbed in the $\bar{K}(j), j = 1, \ldots, \ell$.

## 4.3 Models for Multiple Skills

We have thus far assumed that different types of skills evolve independently. We extend our model to allow vector skills to evolve jointly. We ask whether the improvement in cognitive skills benefits language or motor skills. We also ask if the common scale across levels holds when we consider multiple skill development jointly. Here, we develop the model and a sketch of the proof of identification. We present empirical results for the model in Section 5.

We develop a vector skill formation model, allowing different skill types to evolve jointly. $\ln \boldsymbol{K}(a)$ is a vector of skills at age $a$: [26]

$$\ln \boldsymbol{K}(a) = \boldsymbol{A}' \ln \boldsymbol{K}(a-1) + \boldsymbol{B}' \boldsymbol{\delta}(a)\eta + \boldsymbol{C}' \boldsymbol{V}(Q(a)) + \boldsymbol{\varepsilon}(a). \tag{10}$$

Matrix $\boldsymbol{A}$ captures the transition of current latent skills to next-period skills, and matrix $\boldsymbol{B}$ captures how investments contribute to the skill growth. The term $\boldsymbol{V}(\boldsymbol{Q}(a))$ captures environmental effects growth through maturation and other autogenic effects. $\boldsymbol{\varepsilon}(a)$ is a vector of random shocks at age $a$.

### 4.3.1 Identification of the Multivariate Model

Identification of the model under normal errors follows from application of the analysis of Heckman (1978, 1976). The reduced form (solving $\boldsymbol{K}(a)$ for all inputs up to $a-1$, back to $\boldsymbol{K}(0)$), is in the form of the simultaneous latent variable discrete choice model of Heckman (1978), Case 1. That study draws on the linearity of the

[26] In our model, we consider language, cognitive, and fine motor skills jointly. More details are provided in Appendix L.

system of latent variables and uses standard results in simultaneous equations theory. These results apply to a simultaneous equations latent variables model. The only departure from standard theory is the necessity of making normalizations to the latent variables. We can apply the row–transformation method of Fisher (1966) to secure identification.

To see how to apply his theory, define the set $\mathcal{S} \in \{1, 2, 3\}$ corresponding to the three skills we study. It is straightforward to show that under condition the stated next that there are no admissible row transforms of Equation (10), other than those postulated. The following conditions suffice: (a) independence of the $\varepsilon_\ell(a, s)$ within and across equations and levels, and (b) exogeneity of investment across equations and over time. One normalization is required for each equation, e.g., $\sigma(1, 1, s) = 1$ for each skill $s \in \mathcal{S}$. There are no exclusion restrictions on $\boldsymbol{X}$ across equations, although they vary over ages and levels. $\delta(a, \ell, s)$ is allowed to vary with $s$. For further details, see Appendix L.2.

# 5 Estimates

We use the method of simulated moments to estimate two versions of these models: a) one version allows different skills to develop independently, b) a second version allows skills to develop jointly (vector case). We use more than one thousand moments as our targeted moments. For example, task passing rates for newly enrolled children as the targeted moments for initial conditions; to identify level-specific coefficients, we include each task item passing rate for each difficulty level. For the common scale parameters, we include the covariance of different tasks across adjacent levels. We then report estimates for joint skills. We adjust for clustering in our sample using the paired cluster bootstrap. Details are provided in Appendix I. The moments used in forming the estimates are presented in Table J.1 for the scalar case and in Table L.1 for the vector case. The estimated models pass goodness of fit tests (see Appendix J and L.3). Appendix J also plots model predictions vs. data for each skill, with and without a common scale[27]. In general, imposing the common scale of skill assumption produces worse fits, a point developed further below. The estimates reported in the text do not impose this assumption. Estimates imposing the common scale assumption are presented in Appendix K and L.4. We conduct parallel analysis for scalar and vector cases.

## 5.1 Estimates

We first report empirical results by skill level for the scalar model. We then report results for the vector model. All models allow for discrete measurement errors in the indicator variables measuring knowledge.

### 5.1.1 Language Skills – Scalar Case

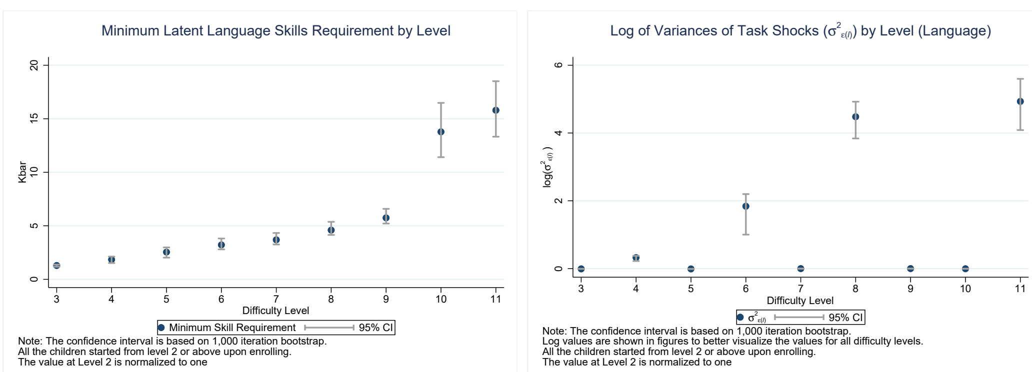
Figure 6a displays estimates of the minimum skill level required at each level. This is defined relative to $\bar{K}(1)$, assuming no shift in model intercepts for each skill across levels apart from that due to skill accumulation. We assume depreciation is not empirically important but can estimate it under the assumption of a common scale

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

of skills. As expected, the skill level required to pass tasks monotonically increases across difficulty levels. We do not impose this 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 6b 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 7, 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 7). Level-specific shocks can intrude to alter the monotonicity delivered by the conventional deterministic model and to capture the lack of fit of the model to the data.[28]

Figure 6: Language Skill Parameters by Difficulty Level

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

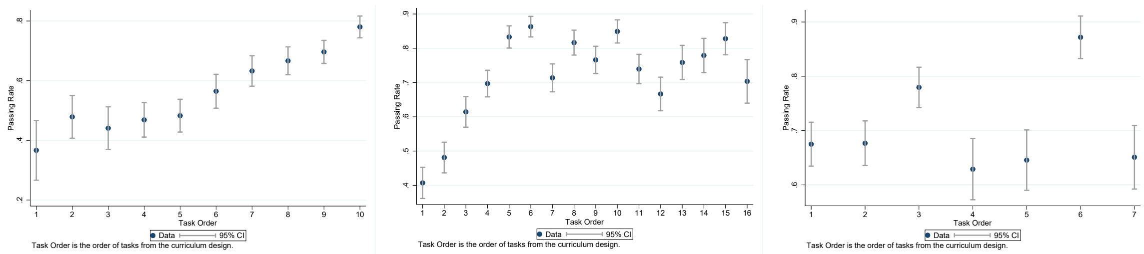



Note that "fadeout" as measured by passing rates, appears within levels 6, 8, and

[28]See Figure J.1b.

Figure 7: Average Passing Rate of Language Tasks by Age Within Level: $p(s, \ell, a)$

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




11 as a consequence of the patterns of item difficulties and variances. This occurs despite the stochastically monotonic increases in skill for all $s$. Variances of shocks differ significantly with levels of skill, across levels, and across skills. See Appendix K for the scalar model and Appendix L.4 for the vector model.

### 5.1.2 Cognitive Skills – Scalar Case

The pattern for the estimated parameters for cognitive skills is similar to that for language skills. For certain difficulty levels, passing rates are not monotone within levels, thus explaining "fadeout" even when, on average, skill levels are increasing.

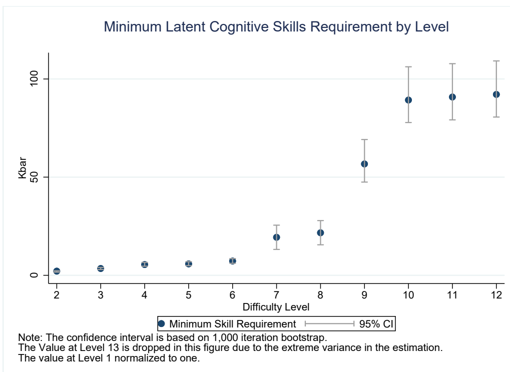
### 5.1.3 Fine Motor Skills – Scalar Case

A similar pattern arises for fine motor skills.

Figure 12 shows how our model can capture the "fadeout" effect. In our model estimates for language variance of task shocks, the variance for level 8 for language skill is large (see Figure 12(a)). The large variance fits the data pattern in Figure 12(b) below. Because the data shows that the children's task performance at level 8 does not monotonically increase and to fit this data pattern, the estimate of

Figure 8: Cognitive Skill

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




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

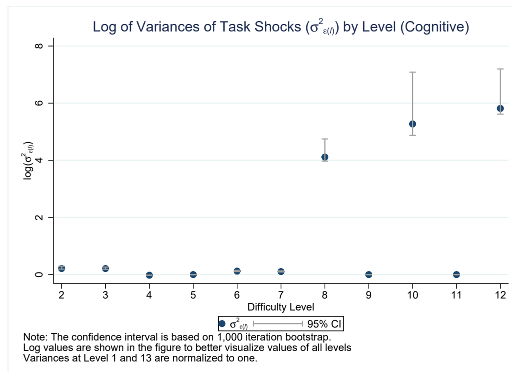
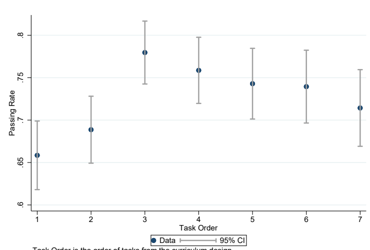
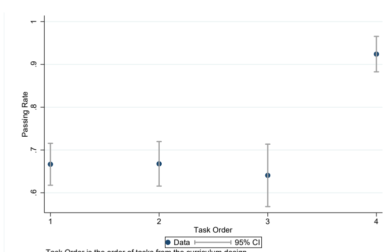



Figure 9: Average Passing Rate of Cognitive Tasks by $a$: $p(s, \ell, a)$

(a) Level 8




(b) Level 10




(c) Level 12

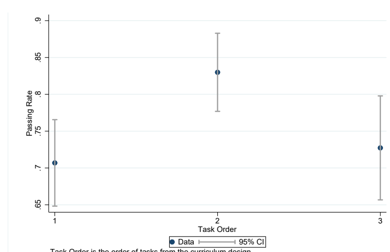


Figure 10: Fine Motor Skill

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

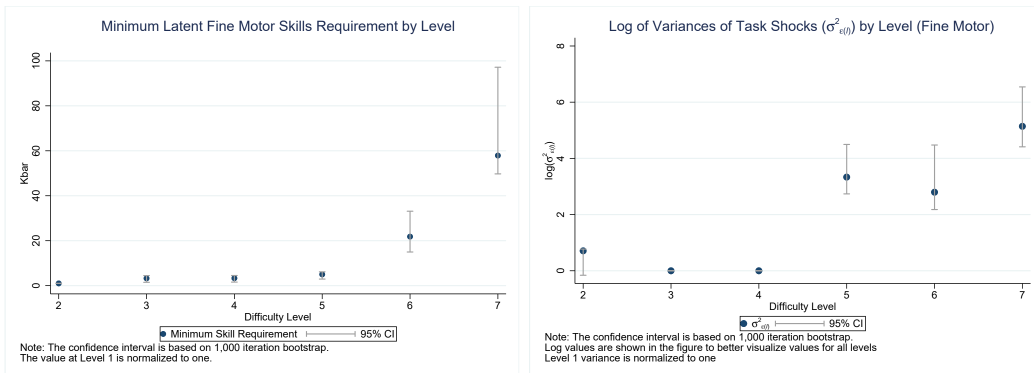
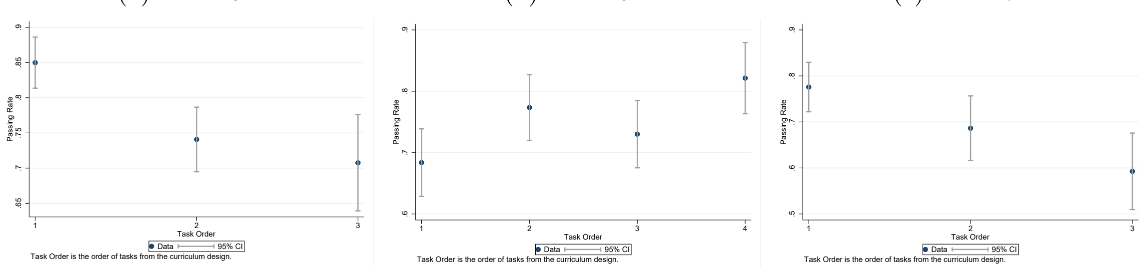



Figure 11: Average Passing Rate of Fine Motor Tasks by Age $a$: $p(s, \ell, a)$

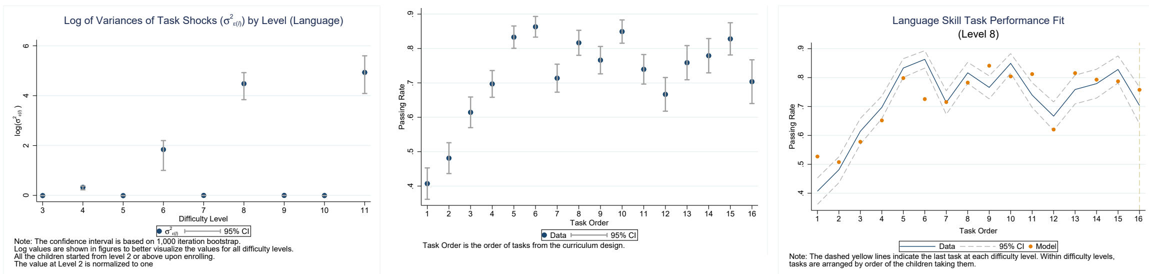
(a) Level 5 (b) Level 6 (c) Level 7



the variance of shock at level 8 has to be large. Figure 12(c) shows that our model fits the data pattern of level 8 very well.

Figure 12: Large Variance to Explain Fadeout

(a) Estimates of $\sigma_{\varepsilon(\ell)}$ (b) Data (c) Model Fit




## 5.2 Learning Components and Task Performance for the Scalar Model

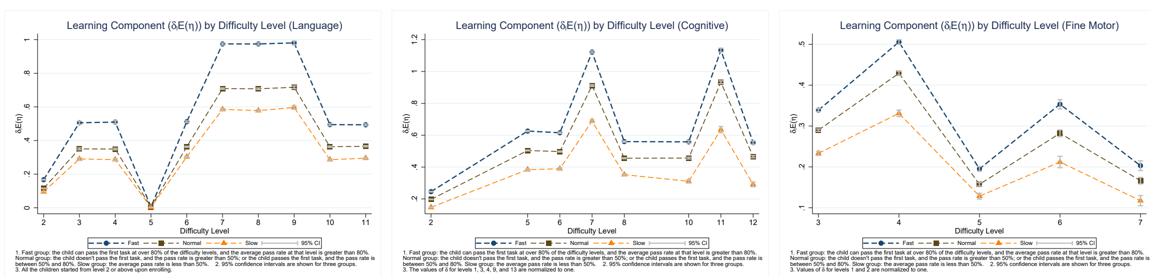
This section examines how the learning component in our structural model $\delta_\ell E(\eta)$ explains child task performance. The $\delta_\ell$ term captures the curriculum content at each difficulty level, which is common across all children (recall tasks within levels have identical learning contexts). The $\eta(\boldsymbol{X})$ term includes interaction quality measures between home visitors and caregivers/children, home visitors' teaching quality, and grandmother rearing during the intervention, and a dummy variable for the gender of the child. These vary with the age of the child and provide important identifying information as noted in Section 4.

The intervention interaction variables (entered as $\boldsymbol{X}$ in $\boldsymbol{\beta_\ell}(\boldsymbol{X})$) are significant determinants of child learning for each task. This finding is consistent with the results in Heckman et al. (2024). The interaction between the home visitor and the caregiver is the only consistently positive interaction that promotes skills (see Appendix K).[29] The grandmother, as the main caregiver, often has significantly negative effects on learning.[30]

Rapid learning (high-ability) children have significantly higher values of the learning component for all skills. This finding is consistent across all difficulty levels for all skills (see Figure 13). 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 K.1). There is learning for children with more educated mothers.

Figure 13: Estimates of $\delta_\ell E(\eta)$ Across Levels by Ability Group for Scalar Models

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




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

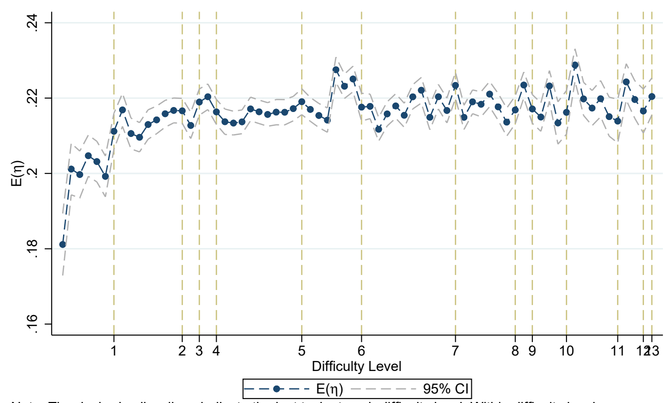
Effects for gender vary by skill. Learning rates are greater for language for girls. For cognitive and fine motor skills, boys learn slightly faster.[31]

[29] All the estimation results are presented in Appendix K.

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

[31] The reported results by gender are for a model without the common scale assumption imposed, a hypothesis we generally reject.

Figure 14a: Learning Component $E(\eta(X))$ of Cognitive Tasks by Level – Scalar Model



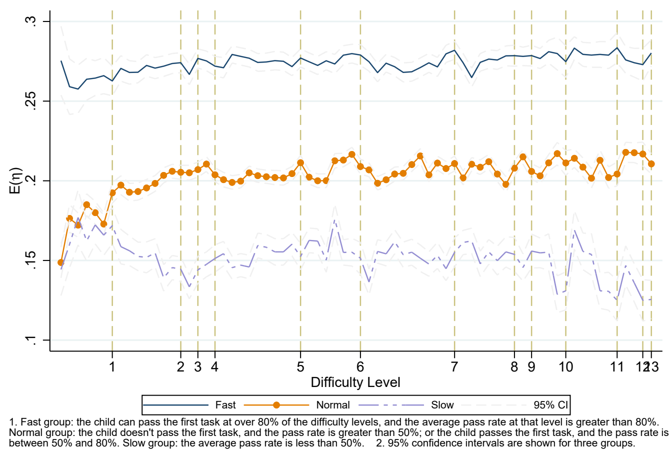

Note: The dashed yellow lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the curriculum design.

We now focus on how the $\eta(\boldsymbol{X})$ term affects child performance on tasks. Figure 14a shows the mean of $\eta(X)$ for each cognitive task. We identify it using $\boldsymbol{\beta_\ell}$ and normalizing $\delta(1) = 1$. In general, there is an increasing pattern of estimated $E(\eta)$ within difficulty levels. In Figure 14b, we break down the estimated $E(\eta)$ values by ability group.[32] 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 estimated learning component $E(\eta)$ increases within a difficulty level, especially strongly for children in the normal group. This finding is also consistent with other skills.[33] For fine motor tasks, there is a similar pattern for tasks

[32]See Table 3 for the definition of the ability groups.
[33]See Figures N.1-N.4 in Appendix N.

greater than 4, although learning is not substantial at any level. For further results, see Appendix N, Figure N.3.

Figure 14b: Learning Component $E(\eta(X))$ of Cognitive Tasks by Level and Ability Group




Appendix Tables N.1-N.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 $\eta$. The interaction quality between the home visitor and the caregiver is higher for households with higher family education levels. Also, the interaction quality measures are significantly different by ability groups and age of enrollment.

## 5.3 Testing for a Common Scale of Skills Across Levels

$\gamma_{1,\ell} = 1$ is consistent with the validity of a common scale of skills connecting $\ell$ and $\ell - 1$. Figure 15 shows that estimates of $\gamma_{1,\ell}$ for each skill level for models estimated without imposing the restriction $\gamma_{1,\ell} = 1$. Table 4 shows the $\chi^2$ test results for each level and skill. Our estimates partially support the common scale assumption. For language and cognitive skills, at some levels, the common scale assumption cannot be rejected. For example, we cannot reject the assumption for language skills between levels 8-11 (i.e., 8-9, 9-10, and 10-11).[34] However, it is decisively rejected in levels 4-6. Table 5 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 a common scale across all levels is rejected. The evidence in favor of a common scale across levels 8-9, 9-10, and 10-11 makes sense, given the similarity of the tasks at those levels. See Table 5. Violations of common scale are also consistent with skill depreciation or appreciation across boundaries.

Figure 15: Tests of the Null Hypothesis of A Common Scale of Skills

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

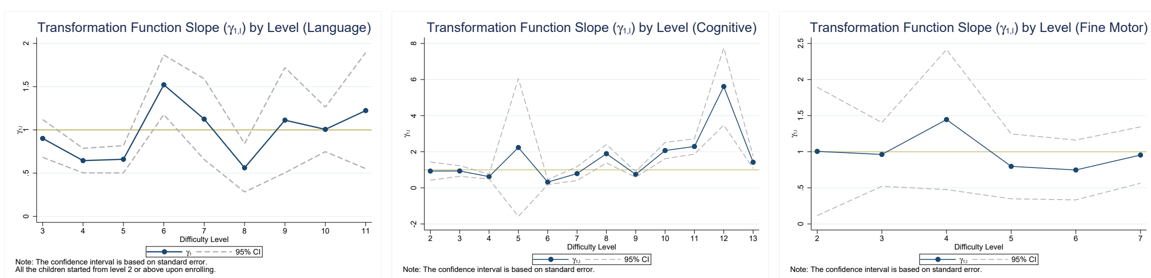



Table 4 reports tests for a common scale for cognitive and fine motor skill tasks. We reject the null of a common scale across virtually all the levels of the cognitive skill tasks. However, we find evidence in support of a common scale for fine motor

[34] $\gamma_{1,\ell} = 1$ implies the existence of a common scale for latent skill variables between level $\ell$ 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.

Table 4: Common Scale Hypothesis Tests by Levels (Scalar Model)

| | 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 5: 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. |

skill tasks, which mainly test drawing skills.

In sum, our estimates do not support the existence of a common scale across most levels for both language and cognitive skills, but the assumption cannot be rejected for some levels and some skills. For example, we cannot reject the common scale assumption for levels 8, 9, and 10 for language skills. It appears to be a valid description of fine motor skills at virtually all levels. Taken as a whole, we think these findings call into question standard practice that relies on an assumed common scale for analyzing skill growth and value-added. Unless one believes that skill depreciation and appreciation operate strongly at the granular level for our estimates, and in the manner just described, we view this as unlikely.

### 5.4 Joint Skill Formation

In Table 6, we report the estimates of matrix $\boldsymbol{A}$. $\boldsymbol{A}_{Cog-lang}$ indicates how the cognitive skill at period $a-1$ contributes to the language skill at period $a$. To simplify the calculations, we impose a common transition matrix across levels.[35] We thus report a summary estimate of $\boldsymbol{A}$.

We find that the diagonal elements are the important ones - the same type of skill is more effective in boosting development. However, skills do not evolve independently. For example, cognitive skills improve both language and fine motor skill development. Language skills improve fine motor skill development, but fine motor skills cannot improve language and cognitive skills. Otherwise, there is little cross-productivity across skills.

[35] In principle, we can tailor the estimates by level, but this leads to a profusion of estimates that are difficult to interpret given the different lesson sequences at the same time across skill levels.

Table 6: Skill Transition Matrix ($\boldsymbol{A}$)

| | | | | | |
|---|---|---|---|---|---|
| $A_{Lang-Lang}$ | 0.933*** | $A_{Lang-Cog}$ | 0.002 | $A_{Lang-Fine}$ | 0.015* |
| | (0.077) | | (0.008) | | (0.009) |
| $A_{Cog-Lang}$ | 0.050** | $A_{Cog-Cog}$ | 0.994*** | $A_{Cog-Fine}$ | 0.038** |
| | (0.020) | | (0.161) | | (0.014) |
| $A_{Fine-Lang}$ | -0.001 | $A_{Fine-Cog}$ | -0.001 | $A_{Fine-Fine}$ | 1.028*** |
| | (0.007) | | (0.008) | | (0.199) |

1. Standard errors are calculated by 500 iteration bootstrap and reported in parentheses.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Estimates of the Investment Transition Matrix ($\boldsymbol{B}$)

| | | | | | |
|---|---|---|---|---|---|
| $B_{Lang-Lang}$ | 0.363*** | $B_{Lang-Cog}$ | 0.001 | $B_{Lang-Fine}$ | 0.014*** |
| | (0.035) | | (0.006) | | (0.006) |
| $B_{Cog-Lang}$ | -0.001 | $B_{Cog-Cog}$ | 1.295*** | $B_{Cog-Fine}$ | 0.015*** |
| | (0.006) | | (0.134) | | (0.006) |
| $B_{Fine-Lang}$ | -0.002 | $B_{Fine-Cog}$ | -0.000 | $B_{Fine-Fine}$ | 1.812*** |
| | (0.007) | | (0.006) | | (0.113) |

1. Standard errors are calculated by 500 iteration bootstrap and reported in parentheses.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We test for cross-productivity of investments across skills, again imposing a common $\boldsymbol{B}$ matrix across levels. Table 7 reports these estimates. Cognitive and language investments enhance the productivity of fine motor skills. Otherwise, there are no other estimated effects that are statistically significant.

We test the common scale assumption in the multivariate model. Table 8 reports test results based on the model described above, allowing different skills to evolve with other skills.[36] The tests for a common scale reported in Table 8 generally support the findings for the scalar model reported in Table 4: the common scale assumption does not hold globally for both language and cognitive skills, but we cannot reject it for fine motor skills.

[36]We report more details in Appendix L.

Table 8: Common Scale Hypothesis Tests by Levels (Vector Model)

| | 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 | | | | 1.070 | 1.235 | 0.267 | 1.066 | 1.814 | 0.178 |
| Level 3 | 1.748 | 7.563 | 0.006 | 0.839 | 6.531 | 0.011 | 1.059 | 0.850 | 0.357 |
| Level 4 | 0.833 | 2.436 | 0.119 | 0.409 | 188.903 | 0.000 | 1.017 | 0.054 | 0.816 |
| Level 5 | 1.332 | 6.231 | 0.013 | 2.816 | 49.930 | 0.000 | 0.967 | 0.473 | 0.492 |
| Level 6 | 1.242 | 6.489 | 0.011 | 0.616 | 135.405 | 0.000 | 0.900 | 7.305 | 0.007 |
| Level 7 | 1.546 | 18.778 | 0.000 | 0.556 | 219.040 | 0.000 | 1.013 | 0.123 | 0.725 |
| Level 8 | 2.007 | 13.910 | 0.000 | 3.555 | 127.810 | 0.000 | | | |
| Level 9 | 1.915 | 3.790 | 0.052 | 0.837 | 2.605 | 0.107 | | | |
| Level 10 | 1.000 | 0.000 | 1.000 | 3.051 | 42.127 | 0.000 | | | |
| Level 11 | 0.551 | 50.794 | 0.000 | 2.912 | 62.423 | 0.000 | | | |
| Level 12 | | | | 8.603 | 932.333 | 0.000 | | | |
| Level 13 | | | | 1.748 | 172.208 | 0.000 | | | |
| | | 109.991 | 0.000 | | 1940.549 | 0.000 | | 10.619 | 0.101 |

## 5.5 Comparing Scalar and Vector Model Estimates

This section compares estimates based on the scalar and vector models. Comparing estimates of the minimum skill requirement $\bar{K}$, we find that for later difficulty levels, the vector model estimates have larger values compared to the estimates from the scalar model (see Figures L.25-L.27 in Appendix L.5.4).

We next compare estimates of the variance of task shocks. We find that estimates from the scalar and the vector models are very close (see Figures L.28-L.30 in Appendix L.5.4). Both models estimate large variances in the task passing rates which do not increase monotonically within the same level. This phenomenon is called "fadeout" in the literature. Similarly, the estimates of $\delta_\ell$, which capture investment components during the intervention, are comparable (see Figures L.31-L.33 in Appendix L.5.4).

## 5.6 How We Account for Measurement Error

In this paper, we follow psychometric conventions (Lord and Novick, 1968) and allow for discrete measurement error when estimating both scalar and vector models. We allow for the possibility that an observation of the child's performance that records a correct answer might come from two sources: a) the child actually knows the task, and b) the child does not know the task but guesses the right answer. For each item of each skill, we allow observations to be recorded with mistakes with the probability $q_s$ for each skill type $s$, which is assumed to be independent across each task given the skill type $s$:

$$\tilde{D}(s,\ell,a) = \begin{cases} D(s,\ell,a) & \text{with probability} \quad 1-q(s) \\ 1-D(s,\ell,a) & \text{with probability} \quad q(s) \end{cases} \tag{11}$$

where $\tilde{D}(s,\ell,a)$ is the recorded value. We allow measurement errors (i.e., home visitors could record by mistake (children passed the task but the record failed or the other way around)).

In Table M.1 in Appendix M, we present the estimates of the probability $q(s)$ for each skill type. Across all difficulty levels, the estimated error probability is not large. Also, given the existence of measurement errors, all estimation results have consistent findings with the model without measurement errors. In a separate analysis, we analyze individual items one-by-one for all skills and estimate very small error probabilities by skill and age (Heckman et al., 2024). See Appendix M, Table M.2 borrows results from that analysis.

# 6 Conclusion

This paper uses novel experimental data on a widely-emulated home visiting program implemented at scale in rural China. We study the mechanisms underlying the positive treatment effects reported in Zhou et al. (2024). The prototypical home visiting intervention we study improves children's skill development through interactions between the home visitor and the caregiver, and not from direct interactions with the child.

Technologies differ across levels of the same skill and across different types of skills. We develop and estimate a latent Markov learning model that captures patterns of learning 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 the decline of measured skills over intervals of time as a consequence of the stochastic nature of learning and the resulting variation in performance across skill assessments.[37] We introduce learning through investment and stochastic shocks into the standard IRT and BKT models of psychometrics.

Girls learn language skills more rapidly than boys. Boys learn cognitive and fine motor skills more rapidly than girls.

We find evidence supporting a common scale of skills across levels for certain skills at certain difficulty levels. However, within our empirically concordant model, we reject the assumption as a global characterization, except for fine motor skills. This finding calls into question the standard practice that assumes the existence of a common scale across levels of scale for analyzing child development across lifetimes

[37] This is sometimes called "fadeout."

and for comparing children.

Cross-fertilization of skills shapes learning, consistent with the evidence in Cunha and Heckman (2008) and Cunha et al. (2010). Cognition promotes the acquisition of both language and fine motor skills. Language skills promote fine motor skills. Fine motor skills have no cross-complementarity effects on other skills. Cognitive and language skill investments bolster the productivity of fine motor investments; otherwise, we find no cross-productivity in investment effects.

This paper uses a concrete measure of investment that consists of educating and motivating the parent. The investment we study is in the *caretaker*. Its impact is mediated by caretaker education levels. Less educated caregivers are less effective vessels of investment. Program designers need to adapt the intervention to bridge the gap between visitor and caregiver when the caregiver is a grandparent, or generally has a lower educational level than the visitor.

The approach taken here enables us to examine the production of skills at a granular level. The technology we estimate departs from the approach that has become standard in the literature in several ways. (1) Scales of skills like those used by Todd and Wolpin (2007), Cunha et al. (2010), and Attanasio et al. (2020) are generally not valid; (2) The technologies for producing skills have qualitatively different characterizations across levels; and (3) If depreciation does operate, it does not operate uniformly across levels of nominally the same skill.

# References

Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2022). When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *Journal of Public Economics 206*, 104574.

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

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: Results from a randomized controlled trial in Colombia. *American Economic Review 110*(1), 48–85.

Bai, Y. (2022). Optimality of matched-pair designs in randomized controlled trials. *Conditionally accepted by the American Economic Review 112*(12), 3911–3940.

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

Bailey, D. H., G. J. Duncan, F. Cunha, B. R. Foorman, and D. S. Yeager (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest 21*(2), 55–97.

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.

Bond, T. and K. Lang (2013). The evolution of the black white test score gap in grades k through 3: The fragility of results. *The Review of Economics and Statistics 95*(5), 1468–1479.

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

Cawley, J., J. J. Heckman, and E. J. Vytlacil (1999, November). On policies to reward the value added by educators. *Review of Economics and Statistics 81*(4), 720–727.

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.

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

Fisher, F. M. (1966, January). *The Identification Problem in Econometrics* (1st Edition ed.). New York: McGraw-Hill.

Freyberger, J. (2022). Normalizations and misspecification in skill formation models. Technical report. arXiv preprint arXiv:2104.00473.

García, J. L. and J. J. Heckman (2023). Parenting promotes social mobility within and across generations. *Annual Review of Economics 15*(1), 349–388.

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. J. (1976). Simultaneous equation models with both continuous and discrete endogenous variables with and without structural shift in the equations. In S. Goldfeld and R. Quandt (Eds.), *Studies in Nonlinear Estimation*, pp. 235–272. Cambridge, MA: Ballinger Publishing Company.

Heckman, J. J. (1978, July). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–959.

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., H. Tian, Z. Zhang, and J. Zhou (2024). Dynamic complementarity: Definitions and nonparametric evidence. Unpublished paper, University of Chicago. Revised from 2023.

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

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

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.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 198–272. Cambridge, MA: MIT Press.

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

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

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

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

Thurstone, L. L. (1927, July). A law of comparative judgement. *Psychological Review 34*(4), 273–286.

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*. University of Illinois Press.

Wachs, T. D., I. C. Uzgiris, and J. M. Hunt (1971). Cognitive development in infants of different age levels and from different environmental backgrounds: An explanatory investigation. *Merrill-Palmer Quarterly of Behavior and Development 17*(4), 283–317.

Zhou, J., J. Heckman, F. Wang, and B. Liu (2023). Early childhood learning patterns for a home visiting program in rural china. *Journal of Community Psychology 51*(2), 584–604.

Zhou, J., J. J. Heckman, B. Liu, and M. Lu (2024). The impacts of a prototypical home visiting program on child skills. *NBER Working Paper No. w27356*. Forthcoming, *Journal of Labor Economics*, January 2026.

Zhou, J., J. J. Heckman, B. Liu, M. Lu, S. M. Chang, and S. Grantham-McGregor (2023). Comparing china REACH and the Jamaica home visiting program. *Pediatrics 151*(Supplement 2).