A Study of the Microdynamics of Early-Childhood Learning

James Heckman

University of Chicago

Jin Zhou

City University of Hong Kong

This paper investigates the weekly evolution of skills as measured by unique data from a widely emulated early-childhood home-visiting program in rural China. The design of the study avoids input endogeneity issues and lack of comparable measures of skills that plague previous studies. Skills, nominally classified as the same, in fact, do not appear to share a common unit scale across levels. They are produced by skill- and life cycle–stage-specific learning processes. A novel dynamic stochastic skill production model for multiple skills is developed, aligning with empirical evidence. The model can explain the "fade-out" and recovery of measures of learning through the operation of a controlled stochastic process.

I. Introduction

This paper uses weekly measurements of skills on children in a prototypical home-visiting program, implemented at scale in China, to investigate

This research was supported in part by National Institutes of Health (NIH) grant NICHD R37HD065072, Research Grants Council (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 NIH and Hong Kong RGC. This paper was presented as a keynote at

Electronically published December 15, 2025

Journal of Political Economy, volume 134, number 1, January 2026.

https://doi.org/10.1086/739256

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 life cycle stages learning occurs, at what rate, and how family environments affect it.

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

$$\overbrace{\boldsymbol{K}(\boldsymbol{a}+1)}^{\substack{\text{skills at}\\ \boldsymbol{a}+1}} = f^{(a)}(\underbrace{\boldsymbol{K}(\boldsymbol{a})}_{\substack{\text{skills at}\\ \boldsymbol{a}}}, \overbrace{\boldsymbol{I}(\boldsymbol{a})}^{\text{investment}}, \underbrace{\boldsymbol{G}(\boldsymbol{a})}_{\substack{\text{environmental}\\ \text{variables}}}). \tag{1}$$

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 by Heckman and Mosso (2014). A recurrent finding of the literature is that enhancements in parenting are associated with improvements in child outcomes (García and

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), Rice University (November 2021), the Cowles conference on Labor and Public Economics (June 2023), City University Hong Kong (June 2024), Society of Economic Dynamics in Barcelona (June 2024), and the Asian Meetings of the Econometric Society in Hangzhou (June 2024). We are grateful for the comments and constructive feedback we received during four rounds of revisions over the past 4 years from four referees. We also extend our sincere thanks to our discussant, Lance Lochner, for his valuable suggestions during the September 2019 Becker Friedman Institute conference in Chicago. Flavio Cunha, Pat Kyllonen, and Tamara McGavock 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. This paper was edited by John List.

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

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, Heckman, and Schennach 2010; Cunha, Nielsen, and Williams 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 (2025) shows the dramatic consequences of different scalings of skill measures for estimates of the technology of skill formation.

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 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 life cycle–stage-specific; the estimated technology differs greatly across levels of nominally the same skill. (2) Investment in caregivers by home

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

[3] See Cunha, Heckman, and Schennach (2010), Bond and Lang (2013), Agostinelli and Wiswall (2021), Cunha, Nielsen, and Williams (2021), and Freyberger (2025).

[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, Heckman, and Schennach (2010), and Attanasio et al. (2020).

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 apparently provide less stimulation in response to the intervention than more educated caregivers.[6] (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 visits received.

The paper unfolds in the following way: Section II describes the background of the program we analyze and its curriculum.[7] Section III presents evidence of learning patterns induced by it. Section IV presents a latent Markov learning model for skills that is concordant with the evidence. Section V presents estimates and interpretations. Section VI concludes.

II. 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 months and 24 months of age. Substantial positive effects are found for the program through age 34 (i.e., Gertler et al. 2014, 2022). Its success has spawned replications around the world, for example, in 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 at scale. Implemented in 2015 by a large-scale randomized controlled trial (RCT), it enrolled 1,500 subjects (from 6 months to 42 months of age) 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, Romano, and Shaikh 2021; Bai 2022). A nonbipartite Mahalanobis matching method was used to pair villages and randomly select one village within a pair into the treatment group and the other

[6] In our sample, grandparent caregivers have, on average, 3 years of education, while parents and home visitors have roughly 10 years of schooling.

Heckman et al. (2025) present a more nuanced nonparametric analysis of this point.

[7] Zhou et al. (forthcoming) describe it in much greater detail.

village into the control group.[8] More details of the design of the experiment and balance tests for treatment and control groups can be found in Zhou et al. (forthcoming).

The intervention cultivates multidimensional skill development through home visiting. Trained home visitors who are roughly at the same level of education as the mothers of the children studied visit each treated household weekly and provide 1 hour of caregiving guidance.

Zhou et al. (forthcoming) 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 end line of the intervention, in contrast with the weekly measurements analyzed here. They 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 difficulty-adjusted 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. (2025) present evidence on dynamic complementarity. The focus of this paper is the growth of skills in the treatment group and not on treatment effects per se.

A. *Enrollment Protocol*

The program enrolls all children ages 6 months to 42 months as of September 2015. Figure 1 gives the enrollment time frame and the Denver assessment timing analyzed by Zhou et al. (forthcoming). 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 and common to all children of the same age. 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.[9] However, the

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

[9] According to the information collected by the China Development Research Foundation field team, 50 villages out of 55 villages have an average 9 years of education for home visitors, which is about 90% of all treated villages. For two villages, home visitors had an average 6 years of education; for two villages, it was about 12 years; and for one village, it was about 14 years. For Pearson's χ^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 same empirical results. Pearson's statistic is $\chi^2(49) = 12.72$ after removing the anomalous villages. We cannot reject

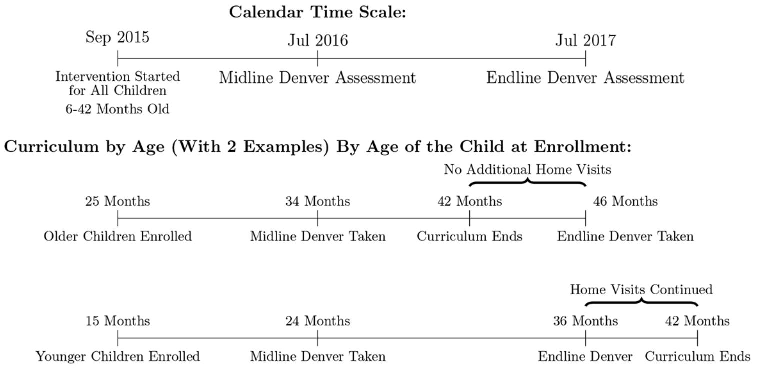

FIG. 1.—China REACH calendar timescales.

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, which is negligible in our sample.

The Denver assessments taken at midline and end line measure child development for both treatment and control children. They are not analyzed in this paper.[10] 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 for two of different age cohorts. The sample cohorts are more or less randomly distributed between 10 months and 25 months old. Table A.1 (tables A.1–N.4 are in the online appendix) documents the balance in backgrounds across different enrollment cohorts. Few children older than 25 months are enrolled.

B. *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.

the null hypothesis that all the villages have the same years of education. Model estimates are essentially the same with and without inclusion of these villages.

[10] They are analyzed in Zhou et al. (forthcoming).

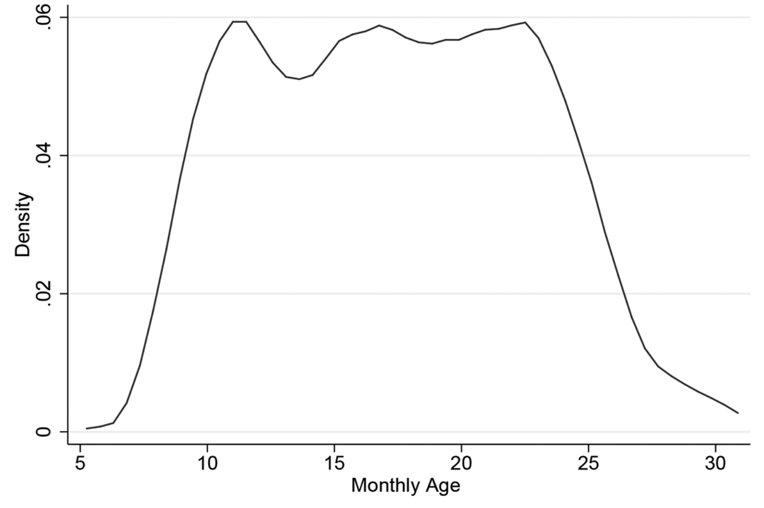

FIG. 2.—Distribution of monthly age when enrolled into program.

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 Uzgiris and Hunt (1975) and Palmer (1971), which we reference in the appendix (available online) as UHP.[11] 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 that promote the skills that appear in these scales.[12] 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.[13] 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. We show the curriculum received for the two hypothetical cohorts in figure 1. The tasks start with simply understanding a picture by verbal acknowledgment to using receptive (heard) language to identify pictures.

[11] More details about the curriculum are provided in appendix sec. B.

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

[13] 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	Cognitive Skill Task
1	Look at pictures and vocalize
2	Name objects and ask baby to point to pictures accordingly
3	Child can name objects in 1 picture and point to named picture
4	Child can name objects in 2 or more pictures and point to named picture
5	Child can point out named pictures and say names of 3 or more
6	Child can point out picture mentioned and correctly name 6 or more pictures
7	Child can talk about pictures, answer questions, understand, or name verbs (eat, play, etc.)
8	Child can follow storyline, name actions, and answer questions
9	Child can understand stories, talk about content in pictures
10	Child can keep up with development of story
11	Child can say name of each graph, discuss role of each item, and then link graphics in card together
12	Child can name things in picture, link different pictures together, and discuss some activities in pictures
13	Child can name things in picture and talk about function of objects

Source.—Scales are from Wachs, Uzgiris, and Hunt (1971).

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. Section C of the appendix gives comparable information for the other skills that follow the same pattern.

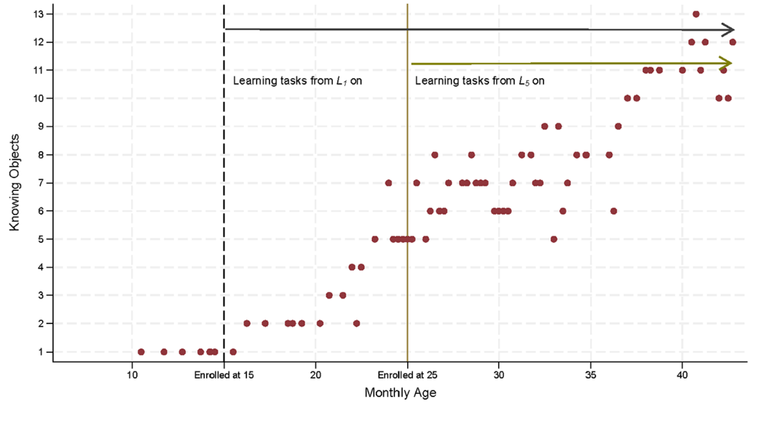

Fig. 3.—Timing of teaching cognitive skills (understand objects): tasks across difficulty levels and two possible enrollment patterns. This figure shows the curriculum for the two hypothetical cases presented in figure 1: an early entrant from level 1 (L_1) and a later entrant from level 5 (L_5). Level 1: look at the pictures and vocalize. Level 13: the child can name things in the pictures and talk about the functions of the objects

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

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

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

III. 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 ℓ and weekly age a, the task item is uniquely determined by 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 ℓ at age a, $K(s, \ell, a)$, which we characterize by

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

where $D(s, \ell, a)$ is the data we observe, recording knowledge of skill s at level ℓ at a given level at age a. Here, $\bar{K}(s, \ell)$ is the minimum-level latent skill required to accomplish the task at difficulty level ℓ. It is the same for all tasks within level ℓ for each s by construction.

This characterization is similar to that used in the classical item response theory (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 ℓ, and let $\bar{a}(s, \ell)$ be the last age at which it is taught at level ℓ. For level ℓ of skill s, indicators of knowledge in a spell are elements of $\{D(s, \ell, a)\}_{\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}$. For example, for cognitive skill level 1, $\underline{a}(s, \ell)$ is age 10 months and 2 weeks, and $\bar{a}(s, \ell)$ is age 15 months and 2 weeks. Seven tasks at level 1 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 ℓ 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. Here, $\Pr(\mathrm{D}(\mathrm{s}, \ell, \bar{\mathrm{a}}(\mathrm{s}, \ell)) = 1)$ is a measure of final skill-s level attainment in level ℓ.

A. *Patterns of Learning*

Figure 4 plots the growth of knowledge in language, cognitive, and fine motor skills.[14] 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.[15] After initial declines, age-specific passing rates within levels increase as learning ensues. The dynamic model presented in section IV 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.[16]

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 section E for a detailed discussion of this measure. Low-ability children learn more slowly.[17]

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

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

[16] In appendix sec. D, we provide the details of the interaction measures.

[17] Heckman et al. (2025) experiment with different definitions of ability and report that their estimates are robust to these alternative measures.

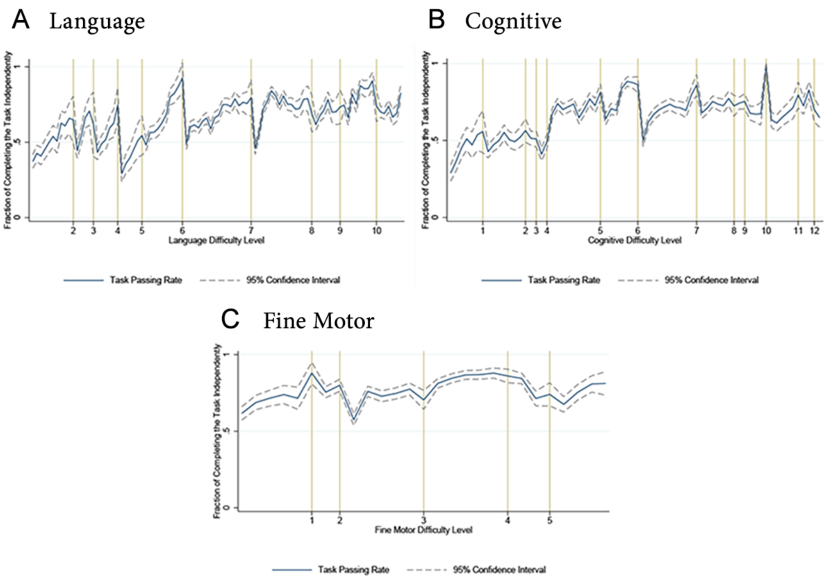

FIG. 4.—Average task passing rates by order and level. *A*, Language. *B*, Cognitive. *C*, Fine motor. The solid vertical lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in the order of the children taking them. *A*, Data are only available at, and beyond, the second level.

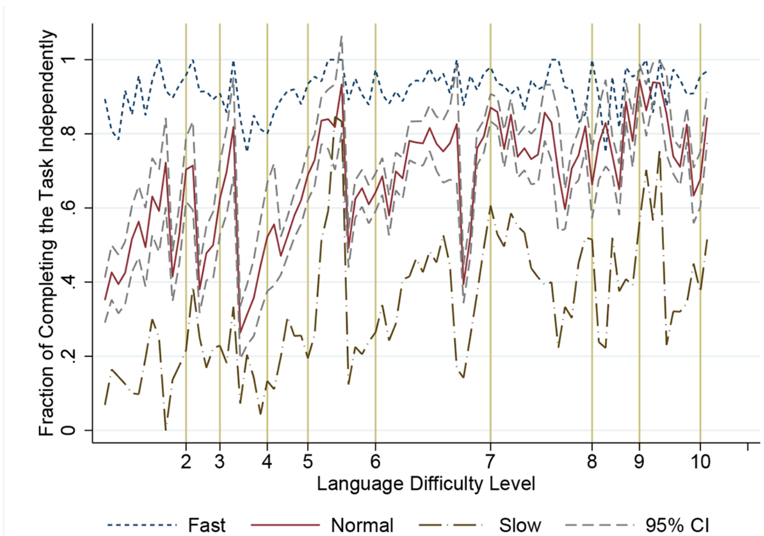

FIG. 5.—By ability group: average language task passing rates.

TABLE 3
Ability Categories (Measured over All Levels)

Group Category	Description
Fast	Pass first task for >80% of difficulty levels, and pass all skill-specific tasks at average rate of >80%
Normal	Pass first task for <80% of difficulty levels, and pass rate is >50%; or pass first task for >80% of difficulty levels, and average passing rate of all skill-specific tasks is 50%–80%
Slow	Average passing rate of all skill-specific tasks is >50%

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 children, on average, have the highest passing rate, a phenomenon that persists across levels and is found for other skills (see figs. F.1–F.4 [figs. A.1–M.4 are in the online appendix]; see also Heckman et al. 2025).

IV. 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 III but suppress the s and ℓ because we initially consider only 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 $K(a) \geq \bar{K}$. If a person at age a masters the skill, $D(a) = 1$, so $D(a) = \mathbf{1}(K(a) \geq \bar{K})$. 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.[18] Skill acquisition is characterized as a random walk:

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

[18] Appendix sec. G compares the empirical performance of multiplicative and additive models. The additive model uses K in place of $\ln K$ in eq. (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.

Here, η 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 external environmental effects that affect the evolution of skills.[19] Further, $V(\boldsymbol{Q}(a))$ is assumed to operate 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 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). We model a controlled random walk.

This economic model extends models in psychometrics—in particular, the IRT model (Lord and Novick 1968)—that measure skills at a point in time. An essential feature of the IRT model is captured by the threshold-crossing feature (2).[20] 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

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

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

[20] The Bayesian knowledge tracing (BKT) model is captured by the dynamics of the model of eq. (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.

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.

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 "fade-out." The literature on fade-out of test scores (see, e.g., Bailey et al. 2020) assumes deterministic growth profiles, whereas we allow for stochastic growth and fade-out of measured skills within a lifetime, accounting for "sleeper effects."[21]

Define $U(a) = \Sigma_{j=1}^{a}\varepsilon(j)$, a random walk, $\Delta(a) = \Sigma_{j=1}^{a}\delta(j)$ is cumulative lessons, and $\Lambda(a) = \Sigma_{j=1}^{a} V(\boldsymbol{Q}(j))$. 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.[22] Conditioning on η, assumed to be independent of $U(a)$ and $K(0)$, we obtain

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

where F_a is the cumulative density function 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 $\mathcal{S}$ 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 ℓ for age a—$\varepsilon_\ell(s, a)$—are assumed to be independent across a. Their distributions may vary with ℓ and s. When estimating the model, we assume that they are i.i.d. within ℓ for each skill s and across s but not necessarily across ℓ, $\eta(s)$ may vary by age a,[23] and $\delta(a)$ captures the content

[21] Sleeper effects are reemergence of a treatment effect for a skill after apparent fade-out.

[22] "Orthogonal" denotes independence.

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

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.

A. *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).[24]

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 (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 appendix section H.4 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 Γ_ℓ is a general function. If there is total depreciation of skills in that transition from $\ell - 1$ to ℓ, Γ_ℓ is the zero function. The property of a common scale across the junction between ℓ and $\ell - 1$ sets $\Gamma_\ell = I$

[24] Cunha, Heckman, and Schennach (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, Heckman, and Schennach (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.

(no depreciation or appreciation) − the identity function. Depreciation of the same skill across junction point ℓ is $\Gamma_\ell = 1 - \sigma_\ell(s)$, where $\sigma_\ell(s)$ is depreciation at level ℓ for skill s. As $\sigma_\ell(s)$ can be negative, there can be appreciation.[25] This paper considers only affine transformations for $\Gamma_\ell(\cdot)$:

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

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.

B. *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 section 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 $Z \perp\!\!\!\perp \Upsilon$. Note $U_1(a) = \Sigma_{j=1}^{a} \varepsilon_1(j)$, 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. We use ω to

[25] This is a one-shot markdown or markup of skill across levels. See Heckman et al. (2025) for evidence on appreciation of skills.

indicate 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 $\ell \in \{1, \dots, L\}$. We can relax this assumption and still achieve identification. However, if we do so, we have to take a position on the dependence across ω_j.[26] 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 $\Sigma_{j=1}^{a} V_1(\boldsymbol{Q}(j))$.

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

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

where $\text{Var}(\Psi_1(a)) = a^2\sigma_\omega^2 + a\sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2 \coloneqq \sigma^2(1, a)$, where $\sigma^2(1, 1) = \sigma_\omega^2 + \sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2$.

Under conditions given in Matzkin (1992, 2007), with sufficient variation in the regressors in period j, $\underline{a}(1) \le j \le \bar{a}(1)$, 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 μ_1 collects any other model intercepts. If any slope coefficient is common across j and j', we can identify the ratio of $\sigma(1, j)/\sigma(1, j')$. Under this condition, with one normalization (e.g., $\sigma(1, j) = 1$), we can identify μ_1^*, $\mu_0(\boldsymbol{Z})$, $\bar{\beta}_1(\boldsymbol{X})$, $V_1(a)$ up to scale. Since we can identify the ratio of $\sigma(1, j)/\sigma(1, j')$, $\sigma(1, a)$, $\sigma(1, a')$ are identified up to a normalization (e.g., a, $a' \neq j$) (see Heckman 1981; 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 $\sigma^2(1, a) \coloneqq a^2\sigma_\omega^2 + a\sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2$, we have the following equations:

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

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 σ_ω^2, $\sigma_{\varepsilon(1)}^2$, and σ_Υ^2. When $a \ge 3$ (i.e., three different tasks at level 1), we can identify all three terms: σ_ω^2, $\sigma_{\varepsilon(1)}^2$, and σ_Υ^2 with sufficient variation in a and j.

[26] 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.

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

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

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

$$\Psi_2(a) = \sum_{j=\underline{a}(2)}^{a} \varepsilon_2(j) + (a - \bar{a}(1))\omega + \sum_{j=1}^{\bar{a}(1)} \varepsilon_1(j) + \bar{a}(1)\omega + \Upsilon,$$

$$\sigma^2(2, a) \coloneqq \mathrm{Var}\Psi_2(a),$$

$$\mathrm{Var}\Psi_2(a) = \sigma^2_{\Upsilon} + a^2\sigma^2_{\omega} + (a - \underline{a}(2))\sigma^2_{\varepsilon(2)} + \bar{a}(1)\sigma^2_{\varepsilon(1)}.$$

Since we have already established identification of σ^2_{ω}, $\sigma^2_{\varepsilon(1)}$, and σ^2_{Υ}, 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 following covariance term:

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

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 $\sigma^2_{\varepsilon(2)}$.[27] 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 ℓ $(\ell \geq 2)$. The only normalization we need is on the scale of variance term $\sigma(1, j) = 1$ at level 1.[28]

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

[28] We can impose any maintained value of $\sigma^{(s)}_{\ell}$.

Under conditions established by Heckman and Vytlacil (2007) and Matzkin (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 section 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.

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

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

From the previous analysis, the term in braces is identified up to the previously stated normalization at the first level, as is the denominator of this expression. Thus $\gamma_{1,2}$ is identified, and we can test whether $\gamma_{1,2} = 1$. 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 ℓ. 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 $\bar{K}(j)$, $j = 1, \ldots, \ell$.

C. *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 whether 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 V.

We develop a vector skill formation model, allowing different skill types to evolve jointly. Here, ln $\boldsymbol{K}(a)$ is a vector of skills at age a:[29]

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

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.

Identification of the multivariate model.—Identification of the model under normal errors follows from the application of the analysis of Heckman (1976, 1978). 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 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 conditions stated next there are no admissible row transforms of equation (10) other than those postulated. The following conditions suffice: (i) independence of the $\varepsilon_\ell(a, s)$ within and across equations and levels and (ii) exogeneity of investment across equations and over time. One normalization is required for each equation, for example, $\sigma(1, 1, s) = 1$ 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 section L.2.

V. Estimates

We use the method of simulated moments to estimate two versions of these models: (a) one version allows different skills to develop independently,

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

and (*b*) a second version allows skills to develop jointly (vector case). We use more than 1,000 moments as our targeted moments. For example, task passing rates for newly enrolled children are 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 section 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 secs. J and L.3). Appendix section J also plots model predictions versus data for each skill, with and without a common scale.[30] 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 sections K and L.4. We conduct parallel analysis for scalar and vector cases.

A. 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.

1. Language Skills—Scalar Case

Figure 6*A* 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 of skills. As expected, the skill level required to pass tasks monotonically increases across difficulty levels. We do not impose this as a restriction on the order of the $\bar{K}(\ell)$ in forming our estimates. 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 6*B* 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 fig. 7). Level-specific shocks can intrude to alter the monotonicity delivered by the

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

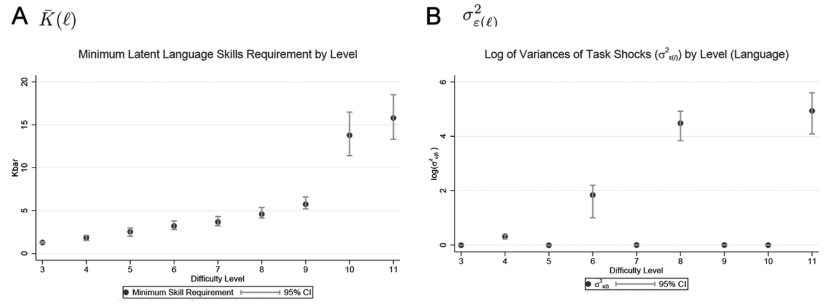

FIG. 6.—Language skill parameters by difficulty level. *A*, $\bar{K}(\ell)$: Minimum latent language skills requirement by level. *B*, $\sigma^2_{\varepsilon(\ell)}$: Log of variances of task shocks to $\sigma^2_{\varepsilon(\ell)}$ by level (language). The confidence interval (CI) is based on a 1,000-iteration bootstrap. All the children started from level 2 or above upon enrolling. The value at level 2 is normalized to one. *B*, Log values are shown in figures to better visualize the values for all difficulty levels.

conventional deterministic model and to capture the lack of fit of the model to the data.[31]

Note that "fade-out" as measured by passing rates, appears within levels 6, 8, and 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 section K for the scalar model and appendix section L.4 for the vector model.

2. Cognitive Skills–Scalar Case

The pattern for the estimated parameters for cognitive skills is similar to that for language skills (see figs. 8, 9). For certain difficulty levels, passing rates are not monotone within levels, thus explaining fade-out even when, on average, skill levels are increasing.

3. Fine Motor Skills–Scalar Case

A similar pattern arises for fine motor skills (see figs. 10, 11).

Figure 12 shows how our model can capture a fade-out effect within our sample. In our model estimates for language variance of task shocks, the variance for level 8 for language skill is large (see fig. 12*A*). The large variance fits the data pattern in figure 12*B* below. Because the data show that the children's task performance at level 8 does not monotonically increase, and to fit this data pattern, the estimate of the variance of

[31] See fig. J.1*b*.

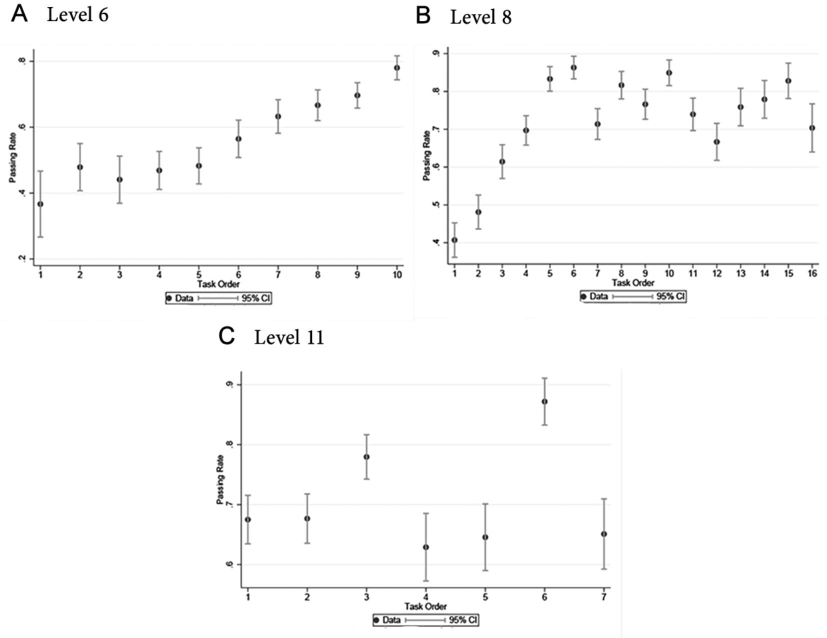

FIG. 7.—Average passing rate of language tasks within level by age a: $p(s, l, a)$. *A*, Level 6. *B*, Level 8. *C*, Level 11. Task order is the order of tasks from the curriculum design.

shock at level 8 has to be large. Figure 12*C* shows that our model fits the data pattern of level 8 very well.

Note that the fade-out discussed here occurs within the sample. Our model might explain the fade-out claimed to exist in the literature. Furthermore, the evidence on fade-out is very weak for high-quality programs. See the discussion in Baulos, García, and Heckman (2025).

B. 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 δ_ℓ 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, plus 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 IV.

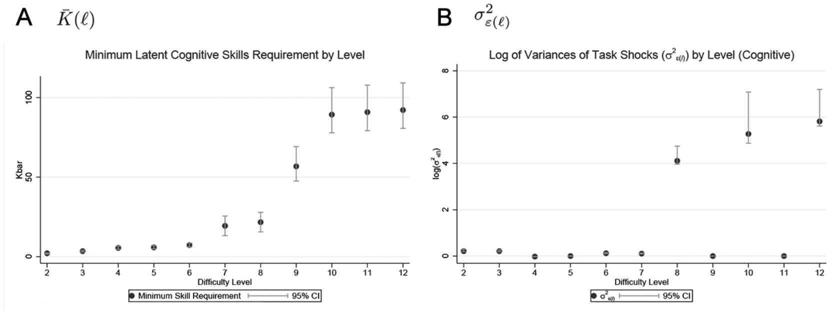

FIG. 8.—Cognitive skill. *A*, $\bar{K}(\ell)$: Minimum latent cognitive skills requirement by level. *B*, $\sigma^2_{\varepsilon(\ell)}$: Log of variances of task shocks to $\sigma^2_{\varepsilon(\ell)}$ by level (cognitive). The confidence interval (CI) is based on a 1,000-iteration bootstrap. *A*, The value at level 13 is dropped in this figure due to the extreme variance in the estimation, and the value at level 1 is normalized to one. *B*, Log values are shown in the figure to better visualize values of all levels; variances at level 1 and level 13 are normalized to one.

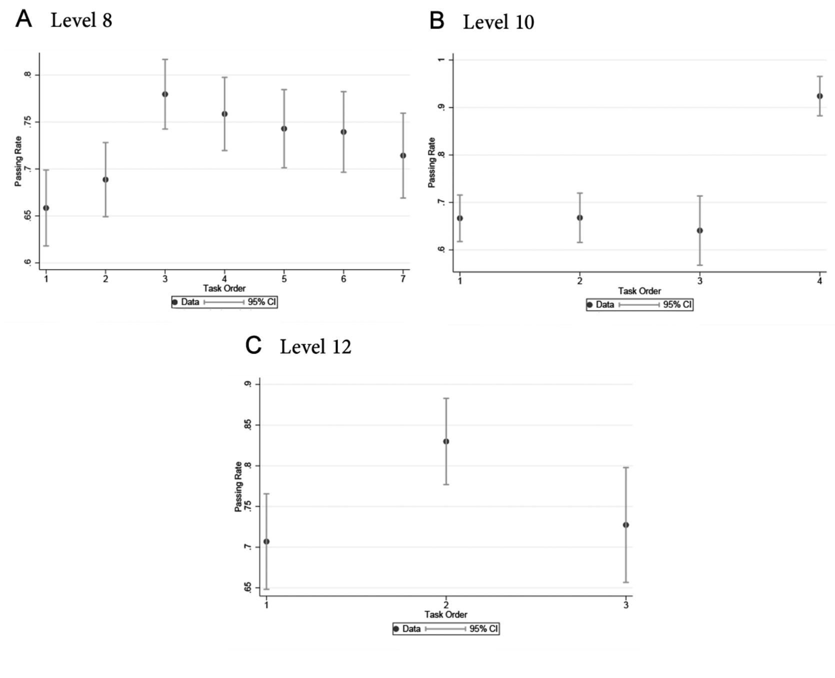

FIG. 9.—Average passing rate of cognitive tasks by age *a*: $p(s, \ell, a)$. *A*, Level 8. *B*, Level 10. *C*, Level 12. Task order is the order of tasks from the curriculum design.

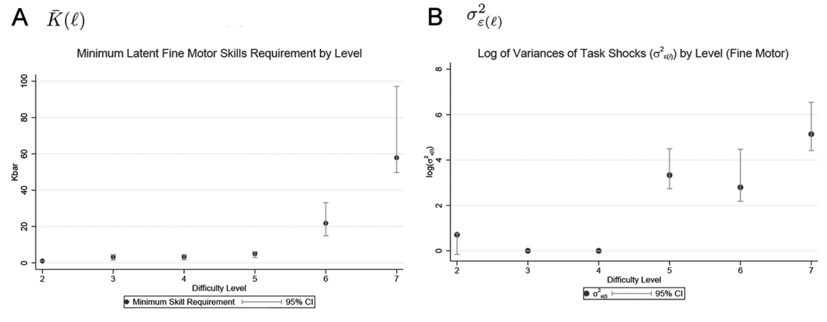

FIG. 10.—Fine motor skill. *A*, $\bar{K}(\ell)$: Minimum latent fine motor skills requirement by level. *B*, $\sigma^2_{\varepsilon(\ell)}$: Log of variances of task shocks $\sigma^2_{\varepsilon(\ell)}$ by level (fine motor). The confidence interval (CI) is based on a 1,000-iteration bootstrap. The value at level 1 is normalized to one. *B*, Log values are shown in the figure to better visualize values for all levels.

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. (2025). The interaction between the home visitor and the caregiver is the only consistently positive interaction

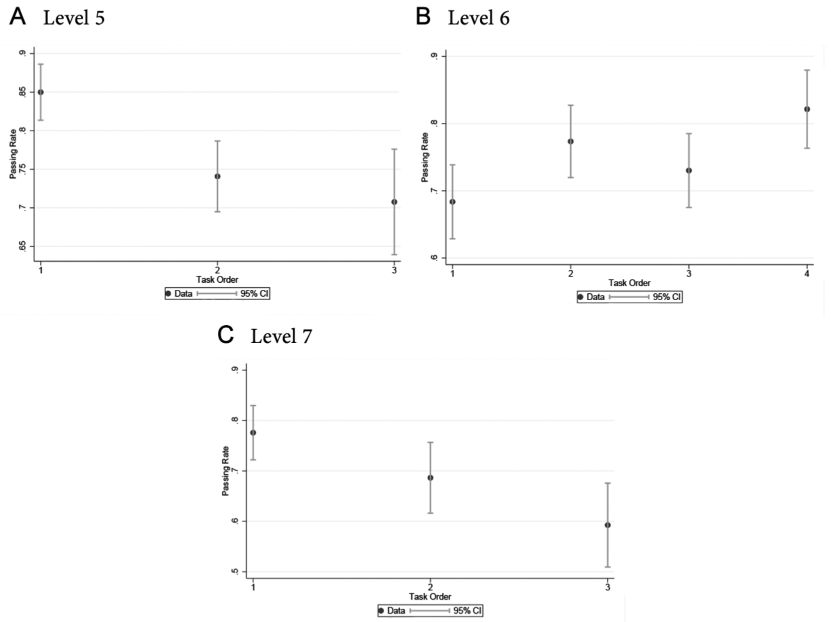

FIG. 11.—Average passing rate of fine motor tasks by age *a*: $p(s, \ell, a)$. *A*, Level 5. *B*, Level 6. *C*, Level 7. Task order is the order of tasks from the curriculum design.

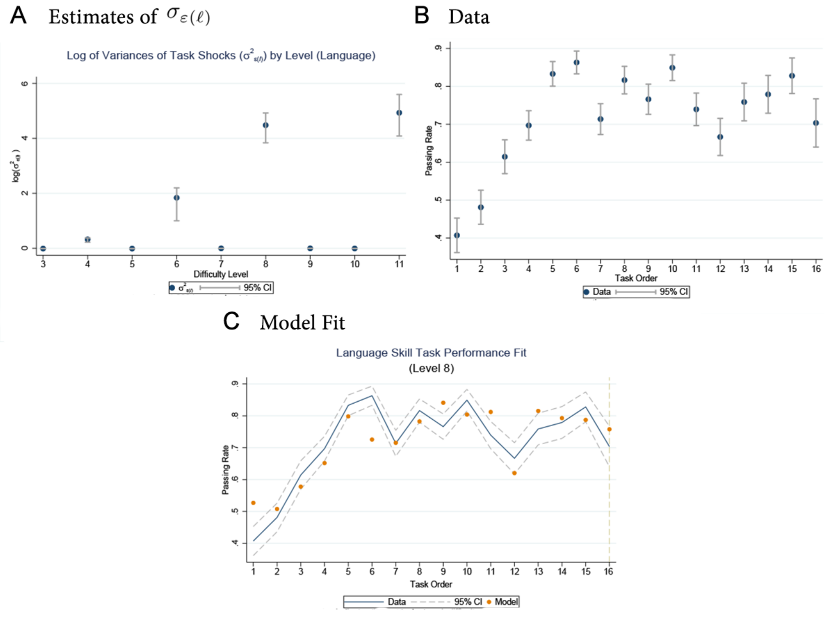

FIG. 12.—Large variances can explain fade-out and recovery. *A*, Log of variances of task shocks $\sigma_{\varepsilon(\ell)}$ by level (language). The confidence interval (CI) is based on a 1,000-iteration bootstrap. Log values are shown in figures to better visualize values for all difficulty levels. All the children started from level 2 or above upon enrolling. The value at level 2 is normalized to one. The value at level 1 is normalized to one. *B*, Data. Task order is the order of tasks from the curriculum design. *C*, Language skill task performance fit (level 8). The dashed vertical lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged by order of the children taking them.

that promotes skills (see appendix sec. K).[32] The grandmother, as the main caregiver, often has significantly negative effects on learning.[33]

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 fig. 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.

Effects for gender vary by skill. Learning rates are greater for language for girls. For cognitive and fine motor skills, boys learn slightly faster.[34]

[32] All the estimation results are presented in appendix sec. K.

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

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

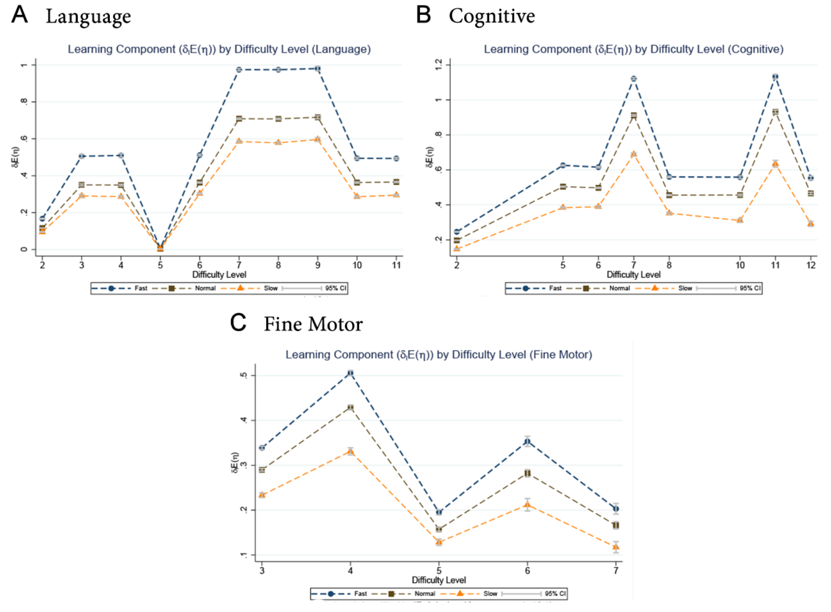

FIG. 13.—Estimates of $\delta_\ell E(\eta)$ across levels by ability group for scalar models. *A*, Learning component $\delta_\ell E(\eta)$ by difficulty level (language). *B*, Learning component $\delta_\ell E(\eta)$ by difficulty level (cognitive). *C*, Learning component $\delta_\ell E(\eta)$ by difficulty level (fine motor). 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 does not 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%. The 95% confidence intervals (CI) are shown for three groups. *A*, All the children started from level 2 or above upon enrolling. *B*, The values of δ for levels 1, 3, 4, 9 and 13 are normalized to one. *C*, The values of δ for levels 1 and 2 are normalized to one.

We now focus on how the $\eta(\boldsymbol{X})$ term affects child performance on tasks. Figure 14*A* 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 14*B*, we break down the estimated $E(\eta)$ values by ability group.[35] 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.

[35] See table 3 for the definition of the ability groups.

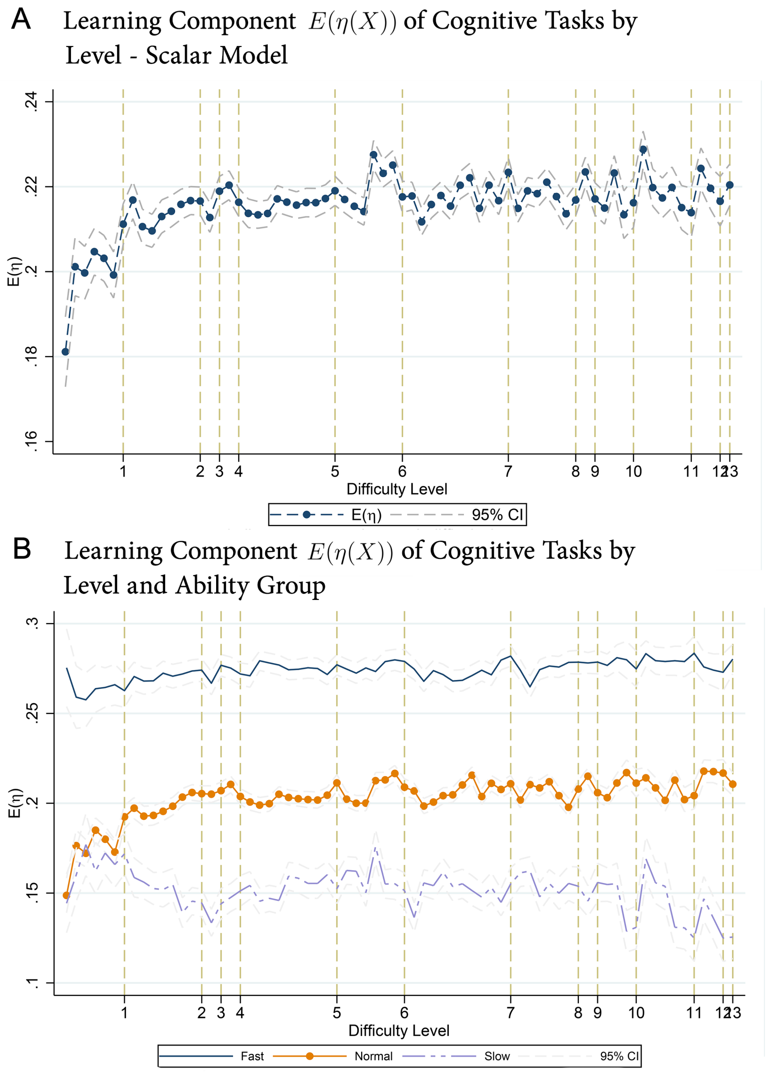

FIG. 14.—*A*, Learning component $E(\eta(X))$ of cognitive tasks by level—scalar model. The dashed vertical lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged in order of the curriculum design. *B*, Learning component $E(\eta(X))$ of cognitive tasks by level and ability group. 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 does not 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%. The 95% confidence intervals (CI) are shown for three groups.

This finding is also consistent with other skills.[36] For fine motor tasks, there is a similar pattern for tasks greater than 4, although learning is not substantial at any level. For further results, see appendix M, figure M.3.

Tables M.1–M.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 households with higher family education levels. Also, the interaction quality measures are significantly different by ability groups and age of enrollment.

C. *Testing for a Common Scale of Skills across Levels*

We show $\gamma_{1,\ell} = 1$ is consistent with the validity of a common scale of skills connecting ℓ 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 χ^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 and 11 (i.e., 8–9, 9–10, and 10–11).[37] 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 the common scale assumption are also consistent with skill depreciation or appreciation across boundaries.

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 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 8–10 for language skills. For example, the coefficient at level 8 indicates that we reject the common scale assumption from level 7 to level 8, and coefficients at levels 9–11 show that we cannot reject the common scale assumption for

[36] See figs. M.1–M.4.

[37] Note $\gamma_{1,\ell} = 1$ implies the existence of a common 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.

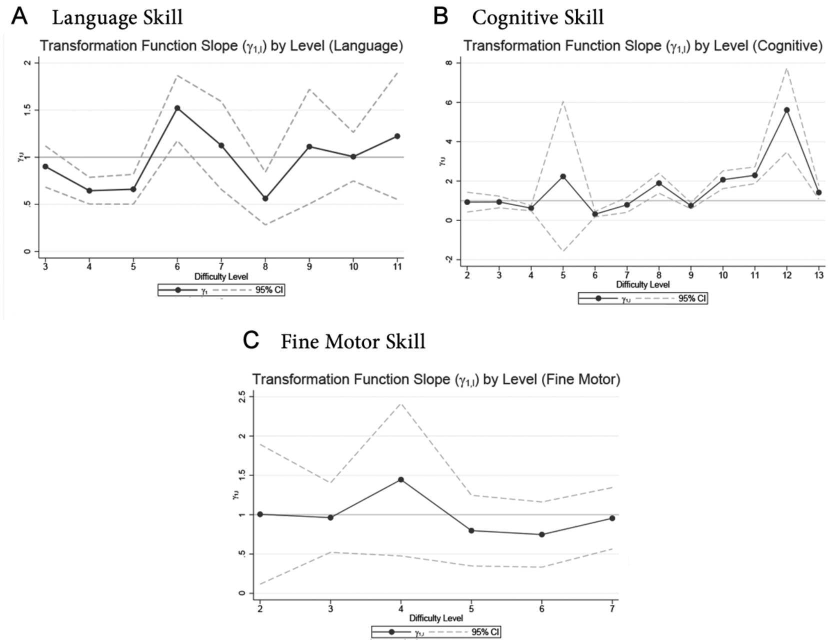

FIG. 15.—Tests of the null hypothesis of a common scale of skills. *A*, Transformation function slope ($\gamma_{1,\ell}$) by level (language). *B*, Transformation function slope ($\gamma_{1,\ell}$) by level (cognitive). *C*, Transformation function slope ($\gamma_{1,\ell}$) by level (fine motor). The confidence interval (CI) is based on standard error. *A*, All the children started from level 2 or above upon enrolling.

levels 8–10 for language skill. 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.

D. *Joint Skill Formation*

In table 6, we report the estimates of matrix $\boldsymbol{A}$: $\boldsymbol{A}_{\text{cog}-\text{lang}}$ 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.[38] 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

[38] 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 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				.929	.012	.914	1.005	.000	.992
Level 3	.901	.546	.460	.936	.010	.922	.963	.022	.883
Level 4	.645	20.193	.000	.621	.142	.707	1.446	.774	.379
Level 5	.66	9.382	.002	2.235	3.899	.048	.798	.720	.396
Level 6	1.522	5.063	.024	.317	17.482	.000	.748	1.277	.258
Level 7	1.125	.182	.670	.791	.362	.547	.955	.034	.853
Level 8	.562	8.195	.004	1.893	4.237	.040			
Level 9	1.113	.113	.737	.744	3.432	.064			
Level 10	1.006	.001	.970	2.068	12.211	.000			
Level 11	1.223	.375	.540	2.292	10.927	.001			
Level 12				5.614	14.351	.000			
Level 13				1.420	4.333	.037			
Total		44.051	.000		71.398	.000		2.827	.830

NOTE.—For each level, we test the null hypothesis that $\gamma_{1,\ell} = 1$. The p-value columns report the probability of not rejecting the null hypothesis. “Total” tests whether the scale invariance assumption is valid across all the levels. Our data for language tasks starts from level 2.

TABLE 5
DIFFICULTY LEVEL LIST FOR LANGUAGE (LEARN WORDS) TASKS

Level	Task
8	Child points to pictures named, names 1 or more pictures, and mimics sound of those objects
9	Child points to pictures named, names 2 or more pictures, and makes sound of those objects
10	Child points at 7 or more pictures and talks about them
11	Teach child some simple descriptive words, and child names objects at home and tells usage of those objects

TABLE 6
SKILL TRANSITION MATRIX ($\boldsymbol{A}$)

$A_{\text{lang-lang}}$	.933***	$A_{\text{lang-cog}}$	.002	$A_{\text{lang-fine}}$	.015*
	(.077)		(.008)		(.009)
$A_{\text{cog-lang}}$	.050**	$A_{\text{cog-cog}}$	.994***	$A_{\text{cog-fine}}$	.038**
	(.020)		(.161)		(.014)
$A_{\text{fine-lang}}$	−.001	$A_{\text{fine-cog}}$	−.001	$A_{\text{fine-fine}}$	1.028***
	(.007)		(.008)		(.199)

NOTE.—Standard errors are calculated by 500-iteration bootstrap and reported in parentheses.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

TABLE 7
ESTIMATES OF THE INVESTMENT TRANSITION MATRIX ($\boldsymbol{B}$)

$B_{\text{lang-lang}}$	.363***	$B_{\text{lang-cog}}$	.001	$B_{\text{lang-fine}}$	.014***
	(.035)		(.006)		(.006)
$B_{\text{cog-lang}}$	−.001	$B_{\text{cog-cog}}$	1.295***	$B_{\text{cog-fine}}$	.015***
	(.006)		(.134)		(.006)
$B_{\text{fine-lang}}$	−.002	$B_{\text{fine-cog}}$	−.000	$B_{\text{fine-fine}}$	1.812***
	(.007)		(.006)		(.113)

NOTE.—Standard errors are calculated by 500-iteration bootstrap and reported in parentheses.
*** $p < .01$.

motor skill development, but fine motor skills do not improve language and cognitive skills. Otherwise, there is little cross-productivity across skills.

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 skills to coevolve with other skills.[39] 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.

E. *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 with the estimates from the scalar model (see figs. L.25–L.27).

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 figs. L.28–L.30). Both models estimate large variances in the task passing rates, which do not increase monotonically within the same level. This phenomenon is called "fade-out" in the literature. Similarly, the estimates of δ_ℓ, which capture investment components during the intervention, are comparable (see figs. L.31–L.33).

F. *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

[39] We report more details in appendix sec. I.

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	.267	1.066	1.814	.178
Level 3	1.748	7.563	.006	.839	6.531	.011	1.059	.850	.357
Level 4	.833	2.436	.119	.409	188.903	.000	1.017	.054	.816
Level 5	1.332	6.231	.013	2.816	49.930	.000	.967	.473	.492
Level 6	1.242	6.489	.011	.616	135.405	.000	.900	7.305	.007
Level 7	1.546	18.778	.000	.556	219.040	.000	1.013	.123	.725
Level 8	2.007	13.910	.000	3.555	127.810	.000			
Level 9	1.915	3.790	.052	.837	2.605	.107			
Level 10	1.000	.000	1.000	3.051	42.127	.000			
Level 11	.551	50.794	.000	2.912	62.423	.000			
Level 12				8.603	932.333	.000			
Level 13				1.748	172.208	.000			
		109.991	.000		1,940.549	.000		10.619	.101

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: (1) the child actually knows the task and (2) 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 for measurement errors; that is, home visitors could record by mistake (children passed the task but the record failed to report it, or the other way around).

In table N.1, 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. 2025). Table N.2 borrows results from that analysis.

VI. 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 by Zhou et al.

(forthcoming). 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.[40] We introduce learning through investment and stochastic shocks into the standard IRT and BKT models of psychometrics.

As reported in the appendix, 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 and for comparing children.

Cross-fertilization of skills shapes learning, consistent with the evidence in Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (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 caretaker. 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, Heckman, and Schennach (2010), and Attanasio et al. (2020) are generally not valid; (2) the technologies for producing skills have qualitatively different

[40] This is sometimes called "fade-out." Recovery and "sleeper effects" are also explained by our models.

characterizations across levels; and (3) if depreciation does operate, it does not operate uniformly across levels of nominally the same skill and it is often followed by appreciation.[41]

Data Availability

Code replicating the results in this article can be found in Heckman and Zhou (2025) in the Harvard Dataverse, https://doi.org/10.7910/DVN/GQBPPO.

References

Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti. 2022. "When the Great Equalizer Shuts Down: Schools, Peers, and Parents in Pandemic Times." *J. Public Econ.* 206:104574.

Agostinelli, F., and M. Wiswall. 2025. "Estimating the Technology of Children's Skill Formation." *J.P.E.* 33 (3): 846–87.

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." *A.E.R.* 110 (1): 48–85.

Bai, Y. 2022. "Optimality of Matched-Pair Designs in Randomized Controlled Trials." *A.E.R.* 112 (12): 3911–40.

Bai, Y., J. P. Romano, and A. M. Shaikh. 2021. "Inference in Experiments with Matched Pairs." *J. American Statis. Assoc.* 117:1726–37.

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 Sci. Public Interest* 21 (2): 55–97.

Baulos, A. W., J. L. García, and J. J. Heckman. 2025. "Perry Preschool at 50: What Lessons Should Be Drawn and Which Criticisms Ignored?" In *The Perry Preschool Study for the 21st Century*, Monographs of the HighScope Educational Research Foundation, no. 15, edited by L. Schweinhart, 88–105. Ypsilanti, MI: HighScope Foundation.

Ben-Porath, Y. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *J.P.E.* 75 (4): 352–65.

Bond, T., and K. Lang. 2013. "The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results." *Rev. Econ. and Statis.* 95 (5): 1468–79.

Bronfenbrenner, U., ed. 2005. *Making Human Beings Human: Bioecological Perspectives on Human Development.* Thousand Oaks, CA: SAGE.

Cawley, J., J. J. Heckman, and E. J. Vytlacil. 1999. "On Policies to Reward the Value Added by Educators." *Rev. Econ. and Statis.* 81 (4): 720–27.

Cunha, F., and J. J. Heckman. 2007. "The Technology of Skill Formation." *A.E.R.* 97 (2): 31–47.

———. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *J. Human Resources* 43 (4): 738–82.

Cunha, F., J. J. Heckman, and S. M. Schennach. 2010. "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." *Ann. Rev. Econ.* 13 (1): 487–513.

[41] See Heckman et al. (2025).

Deonovic, B., M. Yudelson, M. Bolsinova, M. Attali, and G. Maris. 2018. "Learning Meets Assessment." *Behaviormetrika* 45 (2): 457–74.

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

Freyberger, J. 2025. "Normalizations and Misspecification in Skill Formation Models." *Rev. Econ. Studies*, forthcoming.

García, J. L., and J. J. Heckman. 2023. "Parenting Promotes Social Mobility within and across Generations." *Ann. Rev. Econ.* 15 (1): 349–88.

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." Working Paper no. 29292, NBER, Cambridge, MA.

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." *J. Appl. Res. Children Informing Policy Children 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 *Studies in Nonlinear Estimation*, edited by S. Goldfeld and R. Quandt, 235–72. Cambridge, MA: Ballinger.

———. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46 (4): 931–59.

———. 1981. "Statistical Models for Discrete Panel Data." In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. Manski and D. McFadden, 114–78. Cambridge, MA: MIT Press.

Heckman, J. J., and S. Mosso. 2014. "The Economics of Human Development and Social Mobility." *Ann. Rev. Econ.* 6 (1): 689–733.

Heckman, J. J., H. Tian, Z. Zhang, and J. Zhou. 2025. "Dynamic Complementarity." Discussion Paper no. 95/25, Rockwool Foundation, Berlin.

Heckman, J. J., and E. J. Vytlacil. 2007. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics*, vol. 6B, edited by J. J. Heckman and E. E. Leamer, 4779–874. Amsterdam: Elsevier.

Heckman, J. J., and J. Zhou. 2025. "Replication Data for: 'A Study of the Microdynamics of Early Childhood Learning.'" Harvard Dataverse. https://doi.org/10.7910/DVN/GQBPPO.

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, OECD, Paris. 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.

Lu, B., R. Greevy, X. Xu, and C. Beck. 2011. "Optimal Nonbipartite Matching and Its Statistical Applications." *American Statistician* 65 (1): 21–30.

Matzkin, R. L. 1992. "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models." *Econometrica* 60 (2): 239–70.

———. 2007. "Nonparametric Identification." In *Handbook of Econometrics*, vol. 6B, edited by J. J. Heckman and E. E. Leamer. Amsterdam: Elsevier.

McFadden, D. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. Manski and D. McFadden, 198–272. Cambridge, MA: MIT Press.

OECD (Organisation for Economic Co-operation and Development). 2021. *Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills.* Paris: OECD.

Palmer, F. H. 1971. *Concept Training Curriculum for Children Ages Two to Five.* State Univ. New York Stony Brook.

Rutherford, R. S. G. 1955. "Income Distributions: A New Model." *Econometrica* 23 (3): 277–94.

Thelen, E. 2005. "Dynamic Systems Theory and the Complexity of Change." *Psychoanalytic Dialogues* 15:255–83.

Thurstone, L. L. 1927. "A Law of Comparative Judgement." *Psychological Rev.* 34 (4): 273–86.

Todd, P. E., and K. I. Wolpin. 2007. "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps." *J. Human Capital* 1 (1): 91–136.

Uzgiris, I. C., and J. M. Hunt. 1975. *Assessment in Infancy: Ordinal Scales of Psychological Development.* Univ. 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 Q. Behavior and Development* 17 (4): 283–317.

Zhou, J., J. J. Heckman, B. Liu, and M. Lu. Forthcoming. "The Impact of a Prototypical Home Visiting Program on Child Skills." *J. Labor Econ.*, https://www.journals.uchicago.edu/doi/10.1086/732301.

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 (suppl. 2).