A Study of the Microdynamics of Early Childhood Learning

James J. Heckman and Jin Zhou

University of Chicago City University of Hong Kong

Econometric Society in Dynamic Structural Econometrics
HKU 2025

Introduction

Introduction

- The study of early childhood investment and its consequences is important because it is a period of malleability and opportunity in the life of the child
- Featured early childhood interventions have shown significant positive impacts for long-term outcomes (e.g., Perry (center-based) and Jamaica (home visiting) programs)
- Jamaica Reach Up and Learn program has been shown to be low-cost and effective in developing child skills (e.g., Gertler et al., 2022, 2014; Grantham-McGregor and Smith, 2016)

Introduction

- Open questions:
  - What are the key mechanisms through which effective home-visiting interventions produce children's skill development
  - Measuring skills and skill growth:
    There are two notions of skill ***invariance*** in the literature:
    (a) existence of constant units of latent skills
    (b) measurement invariance

    We conduct the test for the first notion: whether there exists constant unit latent skill, a crucial assumption in the literature on skill formation and value-added

Introduction

- Recent literature models the technology of skill formation

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

- Extension of the standard class of **Ben Porath** type models used in the economics of education
- Ben Porath: $\boldsymbol{H}(\boldsymbol{a}+1) = \boldsymbol{g}(\boldsymbol{H}(\boldsymbol{a}), \boldsymbol{I}(\boldsymbol{a})) - \sigma \boldsymbol{H}(\boldsymbol{a})$; scalar H
    - i Age or stage invariant $\boldsymbol{g}(\cdot)$
    - ii Skill assumed to depreciate at rate σ
    - iii Common scale of skill exists over ages

China REACH Program and the Data Analyzed

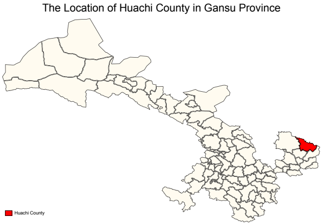
China REACH Program

- China REACH Program applies the Jamaica REACH Up and Learn program, which was pioneered by Sally McGregor and her team in late 1980s.
- China REACH Program provides early childhood development treatments to improve the health and child's (6-42 months) language, cognitive, fine and gross motor skills.
- 2-year randomized controlled trial study in Huachi county.
- The treatment children receive weekly home visits.
- Sample: About 1500 children, who were all age eligible children at baseline in Huachi county.
- Strong treatment effects for skills (see Zhou et al. (2024))
- We just finished the ten years of follow-up in April 2025 and more results are coming!

Huachi County

(a)

(b)



- One of the poorest counties in China: household income is one-third of national-level rural household income.

Huachi County

(c)

(d)

- About 40% of children's fathers work outside of the county. Many mothers migrate or work during the daytime. As a result, in many families (e.g., 30%), grandmothers act as the primary caregiver.
- Among the grandmothers, the average years of education is less than 3 years, and 40% of them do not have any formal education.

Home Visiting Intervention

- Home visitors at the **same** level of education as the mother of the child (e.g., on average 10 years of education and most with preliminary school teaching or nursing experience)
- Trained home visitors visit each treatment household **weekly** and provide **one hour** of parenting or care giving guidance to parent (caregiver) and support based on the adapted Jamaican Reach Up and Learn curriculum
- Supervisors guide home visitors and address problems that home visitors encounter during home visits

China REACH Curriculum

- Well-developed measures UHP (Uzgiris, Hunt, and Palmer, 1970-1975)
- Tasks in the curriculum cover language, cognitive, fine and gross motor skills
- The curriculum is designed based on child weekly age
- In the field, the home visitors strictly follow the design of the curriculum
- Tasks: ordered by difficulty levels
- Within the same level, the tasks with the same content are evaluated multiple times
    - Gives a precise measure of mastery of well-defined tasks and hence, skills
    - Comparable within levels across people
    - Comparability across levels a different matter
    - Gives insight into the details of the skill production literature

China REACH Curriculum

Table 1: Difficulty Level List for Cognitive Skill (Understand Objects) Tasks

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

Source: Uzgiris and Hunt (1975) and Palmer (1971)

China REACH Curriculum

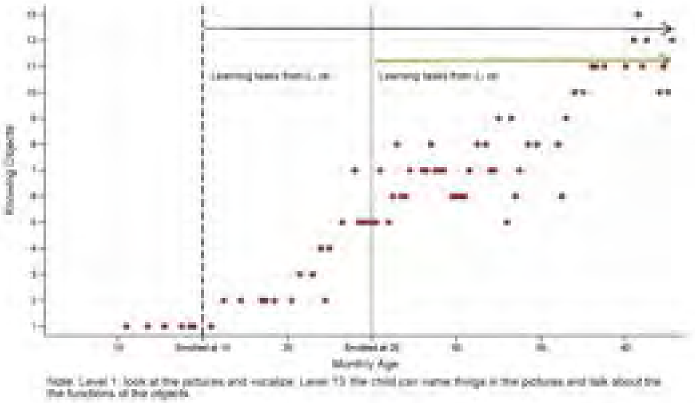
Table 2: Cognitive Skill Task Content: Look at the Pictures and Vocalize (Level 1)

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 makes sound while looking at the pictures

- The same skill is taught within the level

China REACH Curriculum

Figure 1: The Timing of Cognitive Skill (Understand Objects) Tasks across Difficulty Levels



Evidence on Skill Development

Data on Weekly Skill Growth

- $\mathcal{S}$: set of skills taught.
- $\ell(s, a)$: level of skill s taught at age a; $\ell(s, a) \in \{1, \ldots, L_s\}$.
- L_s: number of levels of difficulty for each skill s.
- Mastery (or not) of skill s at level ℓ at age a:

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

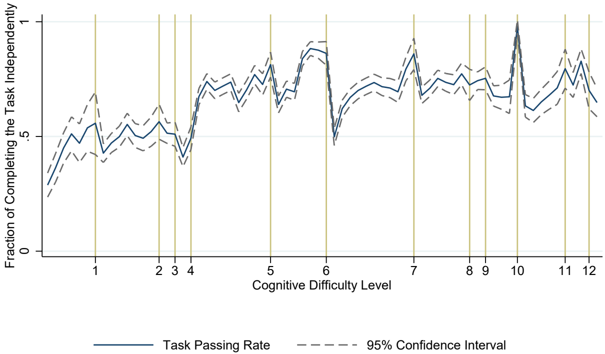
- $\bar{K}(s, \ell)$: minimum latent skill required to master tasks at difficulty level ℓ.
- Indicators of knowledge in a spell:

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

  - $\underline{a}(s, \ell)$: first age at which skill s is taught at level ℓ.
  - $\bar{a}(s, \ell)$: last age at which skill s is taught at level ℓ.

Empirical Finding 1: Test the existence of learning

Figure … d Level



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

- We reject exchangeability and find evidence for knowledge growth, even after controlling for maturation effects.

Link to Robustness (Other Skills)

The Dynamics of Child Learning

Dynamic Learning Model

Start with Single Skill (for only One Difficulty level)

- Program operates at age $a \in [0, \cdots, \bar{A}]$
- Task contents are the same for everyone at age a
- $K(a)$: latent skill at age a
- ***Mastery*** the task at age a at given level if: $K(a) \geq \bar{K}$.
- $D(a) = 1$ if the person at age a masters the skill, so $D(a) = \mathbf{1}\big(K(a) \geq \bar{K}\big)$.

Dynamic Learning Model

- Latent skill evolves within the same difficulty level via

$$\ln \boldsymbol{K}(\boldsymbol{a}) - \ln \boldsymbol{K}(\boldsymbol{a}-1) = \delta(\boldsymbol{a})\eta + \boldsymbol{V}(\boldsymbol{Q}(\boldsymbol{a})) \tag{2}$$

- η: ability to learn and interaction quality measures that are individual-specific and assumed positive ($\eta > 0$).
- $\delta(\boldsymbol{a})$: "lesson" in the curriculum at age $\boldsymbol{a}$. It is the same across all children.
- $\boldsymbol{V}(\boldsymbol{Q}(\boldsymbol{a}))$ captures variables $\boldsymbol{Q}(\boldsymbol{a})$ that affect the evolution of skills that operate independent of the level of $\boldsymbol{K}(\boldsymbol{a}-1)$ and $\delta(\boldsymbol{a})$.
- Since the task content is the same, we assume the unit of latent skill is ***invariant within*** the same difficulty level.

Adding Shocks

- Adding i.i.d. idiosyncratic shocks in growth rates $(\varepsilon(a))$ on a log scale. Skill acquisition is characterized by:

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

Thus

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

$\varepsilon(a)$ i.i.d. across all a and $E(\varepsilon(a)) = 0$.

General Model (Multiple Levels)

- $\mathcal{S}$: set of skills taught
- $\ell(s, a)$: level of skill s taught at age a
- $\ell(s, a) \in \{1, \ldots, L_s\}$
- L_s: number of levels of difficulty for each skill s.

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

- $\bar{K}(s, \ell)$: minimum latent skill requirement to pass the task for difficulty level ℓ.
- $\varepsilon(s, \ell, a)$ i.i.d. across a at level ℓ.

Skill across Difficulty Levels

- The crucial assumption: the common scale of latent skills states that latent skills are comparable across levels as the following equation:

$$\underbrace{K(s,\ell,\underline{a}(s,\ell))}_{\text{Initial condition at level } \ell} = \underbrace{K(s,\ell-1,\bar{a}(s,\ell-1))}_{\text{Terminal condition at level } \ell-1}. \tag{5}$$

- **The above is a measurement relationship at boundary not a production function.**

- Without imposing skill's common scale assumption, the latent skill across difficulty levels evolves via:

$$K(s,\ell,\underline{a}(s,\ell)) = \Gamma_\ell(K(s,\ell-1,\bar{a}(s,\ell-1))) \tag{6}$$

- Under invariance, the Γ_ℓ is identify function (i.e., $\Gamma_\ell(x) = x$).

- In our study, we consider affine transformations as first order approximations:

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

Identification for Difficulty Level One

Allow for heterogeneity:

- Initial condition: $\ln \boldsymbol{K}(0) = \mu_0(\boldsymbol{Z}) + \Upsilon$, $\Upsilon \perp\!\!\!\perp \eta$, and $\boldsymbol{Z} \perp\!\!\!\perp \Upsilon$ (i.e., $\boldsymbol{Z}$ includes family background and the age at the enrollment)
- $\delta_1(\boldsymbol{a})\eta = \bar{\beta}_1(\boldsymbol{X}) + \omega$, $\boldsymbol{X} \perp\!\!\!\perp \varepsilon(1, \boldsymbol{j})$ (i.e., $\boldsymbol{X}$ includes interaction measures and grandmother involvement)
- Rewrite Equation (**??**):

$$\ln \boldsymbol{K}(1, \boldsymbol{a}) = \mu_1 + \mu_0(\boldsymbol{Z}) + \boldsymbol{V}_1(\boldsymbol{a}) + \bar{\beta}_1(\boldsymbol{X})\boldsymbol{a} + \underbrace{\left\{ \boldsymbol{a}\omega + \sum_{\boldsymbol{j}=1}^{\boldsymbol{a}} \varepsilon(1, \boldsymbol{j}) + \Upsilon \right\}}_{\Psi_1(\boldsymbol{a})} \tag{7}$$

- The variance of the sum of shocks at age $\boldsymbol{a}$ at Level 1: $\mathrm{Var}(\Psi_1(\boldsymbol{a})) = \boldsymbol{a}^2\sigma_\omega^2 + \boldsymbol{a}\sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2 := \sigma^2(1, \boldsymbol{a})$.
- $\sigma^2(1, 1) = \sigma_\omega^2 + \sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2$.

Identification for Difficulty Level One

- Latent index model: the key is the identification of the scale of total variance of shocks (e.g., $\sigma(1, \boldsymbol{j})$)
- Under conditions given in Matzkin (1992, 2007), with sufficient variation in the regressors (i.e., $\boldsymbol{X}$, $\boldsymbol{Z}$, and $\boldsymbol{Q}(\boldsymbol{a})$) for this threshold crossing model, we can identify in period $\boldsymbol{j}$, $\underline{\boldsymbol{a}}(1) \leq \boldsymbol{j} \leq \bar{\boldsymbol{a}}(1)$,

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

- Normalize $\sigma(1, 1)$, we can identify μ_1^*, $\mu_0(\boldsymbol{Z})$, $\bar{\beta}_1(\boldsymbol{X})$, $\boldsymbol{V}_1(\boldsymbol{a})$
- If any coefficient is common across $\boldsymbol{j}$ and $\boldsymbol{j}'$, we can identify $\sigma(1, \boldsymbol{j})$
- When $\boldsymbol{j} \geq 3$, can identify σ_ω^2, $\sigma_{\varepsilon(1)}^2$, and σ_Υ^2
  (Notice $\sigma(1, \boldsymbol{j}) = \sqrt{\boldsymbol{j}^2 \sigma_\omega^2 + \boldsymbol{j} \sigma_{\varepsilon(1)}^2 + \sigma_\Upsilon^2}$)

Identification of the Parameters testing Skill's Common Scale Assumption

- Without imposing skill's common scale assumption between level 1 and level 2, we need to discuss how to identify two new parameters $(\gamma_{0,2}, \gamma_{1,2})$:

$$\begin{aligned}\ln \boldsymbol{K}(2, \boldsymbol{a}) &= \mu_2 + \boldsymbol{V}_2(\boldsymbol{a}) + \bar{\beta}_2(\boldsymbol{X})(\boldsymbol{a} - \bar{\boldsymbol{a}}(1)) + \sum_{j=\underline{\boldsymbol{a}}(2)}^{\boldsymbol{a}} \varepsilon(2, \boldsymbol{j}) + \boldsymbol{\Gamma_2}(\ln \boldsymbol{K}(1, \bar{\boldsymbol{a}}(1))) \\ &= \mu_2 + \boldsymbol{V}_2(\boldsymbol{a}) + \bar{\beta}_2(\boldsymbol{X})(\boldsymbol{a} - \bar{\boldsymbol{a}}(1)) + \sum_{j=\underline{\boldsymbol{a}}(2)}^{\boldsymbol{a}} \varepsilon(2, \boldsymbol{j}) \\ &\quad + (\boldsymbol{\gamma_{0,2}} + \boldsymbol{\gamma_{1,2}} \ln \boldsymbol{K}(1, \bar{\boldsymbol{a}}(1))).\end{aligned}$$

- Using the covariance across levels, we can identify the parameter $\gamma_{1,2}$:

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

- The key is to test $\gamma_{1,2} = 1$

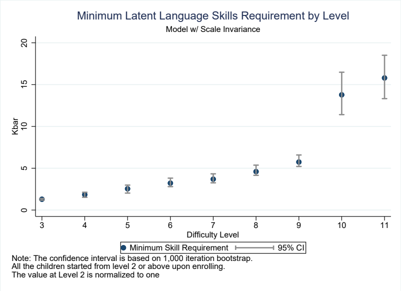
Link to Higher Level

Estimation Results

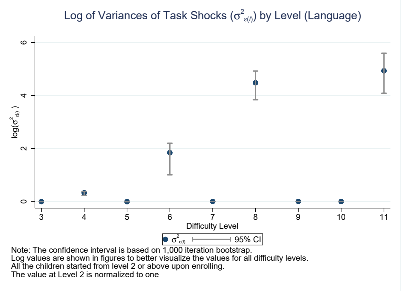
Language Skills

Figure 3: Language Skill

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



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

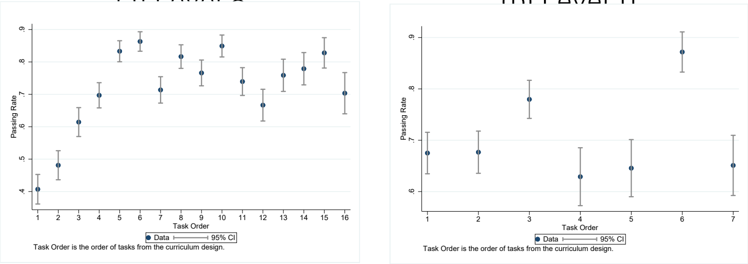


- Figure 3(a) displays estimates of the estimated minimum skill requirement for each level.
- The variances at levels 8, and 11 are larger than the variances at other levels.

Language Skills

Figure 4: Average Passing Rate of Language Tasks: $p(s, \ell)$ (Raw Data)

(a) Level 8 (b) Level 11

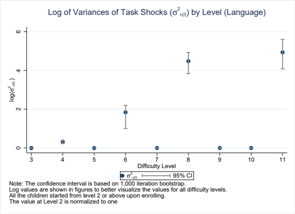


- Our model can explain the “fadeout” measured by passing rate using variances of task specific shocks by difficulty levels

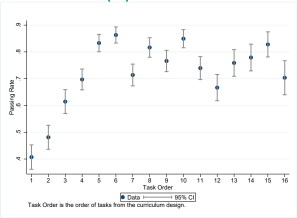
Language Skills

Figure 5: Large Variance to Explain Fadeout

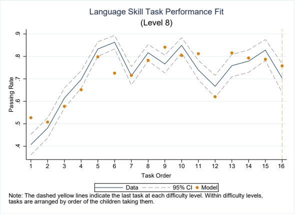
(a) Estimates of $\sigma_{\varepsilon(\ell)}$



(b) Data



(c) Model Fit



- 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 the variance of shock at level 8 has to be large.
- Figure 5(c) shows that our model fits the data pattern of level 8 very well.

Testing Common Scale ($\gamma_{1,\ell} = 1$)

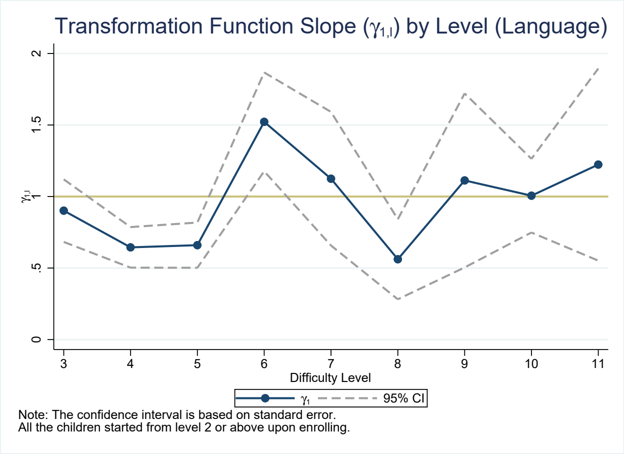
Table 3: Common Scale Hypothesis Tests by Levels ($\gamma_{1,\ell} = 1$)

	Language		Cognitive		Fine Motor	
	Slope($\gamma_{1,\ell}$)	p-value	Slope($\gamma_{1,\ell}$)	p-value	Slope($\gamma_{1,\ell}$)	p-value
Level 2			0.929	0.914	1.005	0.992
Level 3	0.901	0.460	0.936	0.922	0.963	0.883
Level 4	0.645	0.000	0.621	0.707	1.446	0.379
Level 5	0.660	0.002	2.235	0.048	0.798	0.396
Level 6	1.522	0.024	0.317	0.000	0.748	0.258
Level 7	1.125	0.670	0.791	0.547	0.955	0.853
Level 8	0.562	0.004	1.893	0.040		
Level 9	1.113	0.737	0.744	0.064		
Level 10	1.006	0.970	2.068	0.000		
Level 11	1.223	0.540	2.292	0.001		
Level 12			5.614	0.000		
Level 13			1.420	0.037		
Total $\chi^2(\cdot)$	44.051	0.000	71.398	0.000	2.827	0.830

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

Testing Common Scale

Figure 6: Tests of the Null Hypothesis of Common Scale acorss Levels



Link to Other Skills

Testing Common Scale

Table 4: Language Tasks for Difficulty Levels 8-11

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.

Cross-Fertilization and Multiple Skill Development

Joint Skill Formation

- Vector version of skill formation, allowing different skill types to evolve jointly

$$\ln \mathbf{K}(a) = \mathbf{A}' \ln \mathbf{K}(a-1) + \mathbf{B}'\boldsymbol{\delta}(a)\eta + \mathbf{C}'\mathbf{V}(Q(a)) + \varepsilon(a). \quad (8)$$

- The matrix $\mathbf{A}$ captures the transition of current latent skills to next-period skills
- The matrix $\mathbf{B}$ captures how investments contribute to the skill growth
- The term $\mathbf{V}(\mathbf{Q}(a))$ captures environmental effects growth through maturation and other autogenic effects.
- Identification follows from previous proof due to exogeneity of inputs from design of intervention

Joint Skill Formation

Table 5: Skill Transition Matrix (**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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

- We find that the diagonal elements are the important ones - the same type of skill is more effective in boosting development.
- Skills do not evolve independently.
  - Cognitive skills improve both language and fine motor skill development
  - Language skills improve fine motor skill development
  - Fine motor skill cannot improve language and cognitive skill

Joint Skill Formation

Table 6: 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

Conclusion

1. Home visiting intervention improves children's skill development through the key mechanism: interaction between the home visitor and the caregiver
2. Evidence consistent with dynamic complementarity using nonparametric methods (Critical or sensitive periods).
3. An empirically concordant dynamic model of learning with a generalized setting
4. Evidence supporting skill's common scale for certain skills at specific difficulty levels, but not globally for most skills
5. Cross fertilization is an important component of learning
6. Technology differs across levels of nominally the same skill

Appendix

D. Measures of Interactions and Factor Model of Interaction

D.1 Measures of Interactions

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

Table 21: Measures Used for Interaction Quality

Between Home Visitor and Caregiver
Has the home visitor explained the task content and lesson target to the caregiver?
Has the home visitor shown the lessons and given examples to the caregiver?
Does the home visitor ask the caregiver to play the lessons with the child alone?
Does the caregiver ask the home visitor about lessons in the next week?
Has the home visitor listened to the caregiver?
Has the home visitor answered the caregiver's questions?
Has the home visitor asked for the caregiver's opinions?
Does the home visitor encourage and help the caregiver?
Is the relationship between the home visitor and caregiver friendly, understandable, and cooperated?
Has the home visitor discussed with a caregiver or other persons about the content not related to the home visiting?

Table 21: Measures Used for Interaction Quality, Cont'd

Between Home Visitor and Child
Has the home visitor shown the lessons and given examples to the child?
Has the home visitor explained the lesson to the child?
Does the home visitor listen to the child and respond to the child's voice or action?
Does the home visitor praise the child when the child tries to master one task?
Does the home visitor use language to communicate with the child when the child is completing the lessons?
Does the home visitor give the child enough time to explore the materials and finish the lessons?
Is the relationship between the home visitor, and the child-friendly, understandable, and cooperative?

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

Table 22: Measures Used for Teaching Quality

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

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

Return to main text

D.2 Factor Model for Summarizing Interactions

- Let $\boldsymbol{M}_{ia}^{j,l}$ be the measure $\boldsymbol{j}$ at household $\boldsymbol{i}$ at the child's age $\boldsymbol{a}$ and γ_{ia}^{l} be the latent factors $\boldsymbol{l}$ (i.e., teaching ability, the interaction quality between home visitor and caregiver, and the interaction quality between home visitor and child).

$$\boldsymbol{M}_{ia}^{j,l} = \boldsymbol{X}_{ia}'\beta + \alpha^{j}\gamma_{ia}^{l} + \epsilon_{ia}^{j,l}.$$

- Since the measure $\boldsymbol{M}_{ia}$ is a categorical variable, we use ordered probit model with latent factor γ and estimate the factor model by MLE assuming ϵ_{ia} is from normal distribution with zero mean.

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

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

- Therefore, the latent factor estimate for $\boldsymbol{l}$ is

$$\hat{\gamma}^l = \int \gamma \boldsymbol{g}(\gamma|\boldsymbol{M}^l,\boldsymbol{X};\beta,\alpha)\boldsymbol{d}\gamma^l.$$

- The prior distribution of ϕ is based on an estimated factor distribution.

Table 23: Prior Variances for Latent Factors

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

Table 24: Factor Model: Teacher Ability

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

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

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

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

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

Table 25: Factor Model: Interactions Quality between Home Visitor and Caregiver

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

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

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

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

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

Table 26: Factor Model: Interactions Quality between Home Visitor and Child

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

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

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

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

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

Return to main text

H. Identification

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

$$
\begin{aligned}
\ln \boldsymbol{K}(\ell, \boldsymbol{a}) &= \ln \boldsymbol{K}(\ell, \boldsymbol{a} - 1) + \delta_\ell(\boldsymbol{a})\eta + \boldsymbol{V}_\ell(\boldsymbol{Q}(\boldsymbol{a})) + \varepsilon(\ell, \boldsymbol{a}), \\
&\qquad \underline{\boldsymbol{a}}(\ell) \leq \boldsymbol{a} \leq \bar{\boldsymbol{a}}(\ell). \\
\boldsymbol{E}(\varepsilon(\ell, \boldsymbol{a})) &= 0. \\
\mathrm{Var}(\varepsilon(\ell, \boldsymbol{a})) &= \sigma^2_{\varepsilon(\ell)}.
\end{aligned}
$$

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

$$\ln \boldsymbol{K}(\ell, \underline{\boldsymbol{a}}(\ell)) = \ln \boldsymbol{K}(\ell - 1, \bar{\boldsymbol{a}}(\ell - 1)) + \delta_\ell(\underline{\boldsymbol{a}}(\ell))\eta + \boldsymbol{V}_\ell(\boldsymbol{Q}(\underline{\boldsymbol{a}}(\ell))) \\ + \varepsilon_\ell(\boldsymbol{a}(\ell)), \qquad \underline{\boldsymbol{a}}(\ell) \leq \boldsymbol{a} \leq \bar{\boldsymbol{a}}(\ell).$$

- Define

$$\Delta_\ell(\boldsymbol{a}) := \sum_{\boldsymbol{j}=\underline{\boldsymbol{a}}(\ell)}^{\boldsymbol{a}} \delta_\ell(\boldsymbol{j}).$$

$$\Lambda_\ell(\boldsymbol{a}) := \sum_{\boldsymbol{k}=\underline{\boldsymbol{a}}(\ell)}^{\boldsymbol{a}} \boldsymbol{V}_\ell(\boldsymbol{Q}(\boldsymbol{k})).$$

$$\boldsymbol{U}_\ell(\boldsymbol{a}) := \sum_{\boldsymbol{j}=\underline{\boldsymbol{a}}(\ell)}^{\boldsymbol{a}} \varepsilon(\ell, \boldsymbol{j}).$$

- We assume $\delta_\ell(\boldsymbol{a}) = \delta_\ell$ (for all $\boldsymbol{a}$ in ℓ) is consistent with the same skill being taught at each level.
- We parameterize the model such that

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

  - where $\boldsymbol{X}$ includes determinants of ability and learning such as family background and interaction measures.

- In this notation, $\eta\Delta_\ell(\boldsymbol{a}) := (\boldsymbol{a} - \underline{\boldsymbol{a}}(\ell))(\bar{\beta}_\ell(\boldsymbol{X}) + \omega_\ell)$.
- We assume that $\omega_\ell = \omega$ for all ℓ, $\boldsymbol{E}(\omega) = 0$, and $\mathrm{Var}(\omega) = \sigma_\omega^2$.
- We assume normal errors in making our estimates. Thus, we restrict attention to identification of means and covariances.

H.1 Recursive Definition of the Skill Index

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

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

- It is useful to collect the unobservables into

$$\Psi_1(\boldsymbol{a}) = \{\boldsymbol{a}\omega + \boldsymbol{U}_1(\boldsymbol{a}) + \Upsilon\}.$$

- We assume (ω, Υ) are mutually independent, that $\varepsilon_1(\boldsymbol{a})$ is independent of $\varepsilon_1(\boldsymbol{a}')$, and $(\omega, \Upsilon) \perp\!\!\!\perp \varepsilon_1(\boldsymbol{a})$ for all $\boldsymbol{a}$.

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

- We define $\mathrm{Var}(\Psi(\boldsymbol{a})) := \sigma^2(1, \boldsymbol{a})$.

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

H.2 Identification of First Level Parameters

- Under conditions presented in Matzkin (1993, 2007), we can identify up to scale $\sigma(1,1)$.
- This follows from standard results in the binary choice model, provided that at least one parameter is constant over the ℓ interval.
- In estimation, we assume $\boldsymbol{V}_\ell(\Lambda_\ell(\boldsymbol{a})) := \sum_{j=0}^{J} \lambda_j \boldsymbol{a}^j$ to capture aging and maturation effects, which we assume operates uniformly across ℓ.
- We assume $\lambda_j = 0, \boldsymbol{j} > 2$.
- Although in principle we could identify ℓ-specific maturation effects, we do not do it here.
- Shocks are i.i.d. with mean zero (i.e., $\boldsymbol{E}(\varepsilon_\ell(\boldsymbol{a})) = 0$) and with variance $\sigma^2_{\varepsilon(\ell)}$.

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

$$\ln \boldsymbol{K}(\ell, \boldsymbol{a}) = \eta \Delta_\ell(\boldsymbol{a}) + \Lambda_\ell(\boldsymbol{a}) + \boldsymbol{U}_\ell(\boldsymbol{a}) + \ln \boldsymbol{K}(\ell, \underline{\boldsymbol{a}}(\ell)). \tag{9}$$

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

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

$$\begin{aligned} \Pr(\boldsymbol{D}(\ell, \boldsymbol{a}) = 1) &= \Pr(\ln \boldsymbol{K}(\ell, \boldsymbol{a}) \geq \ln \bar{\boldsymbol{K}}_\ell) \\ &= \Pr(\ln \boldsymbol{K}(\ell, \underline{\boldsymbol{a}}(\ell)) + \eta \Delta_\ell(\boldsymbol{a}) + \Lambda_\ell(\boldsymbol{a}) + \boldsymbol{U}_\ell(\boldsymbol{a}) \geq \ln \bar{\boldsymbol{K}}_\ell). \end{aligned}$$

- The introduction of $\bar{\boldsymbol{K}}_\ell$ in the estimation adds an intercept to the model that is the same for each $\boldsymbol{\alpha}$ in ℓ, but in general differs across $\boldsymbol{s}$.
- For $\ell = 1$, we characterize the initial condition by $\ln \boldsymbol{K}(1, \underline{\boldsymbol{\alpha}}(1)) = \mu_0(\boldsymbol{Z}) + \Upsilon$, $\boldsymbol{E}(\Upsilon) = 0$, $\mathrm{Var}(\Upsilon) = \sigma^2_\Upsilon$, for the interval $[\underline{\boldsymbol{\alpha}}(1), \bar{\boldsymbol{\alpha}}(1)]$.
- Thus, at $\underline{\boldsymbol{\alpha}}(1) = 1$, we can identify

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

- μ_1 is the intercept which includes $\bar{\boldsymbol{K}}_1$.

- At the next age, we can identify

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

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

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

- Using the same logic across ages for $\ell = 1$, we can identify

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

- We can identify $\bar{\boldsymbol{K}}_\ell$, $\ell = 1, \ldots, \boldsymbol{L}$ up to scale and an unknown constant if we assume intercepts are constant across levels.
- Once we normalize $\sigma(1, 1)$, we can identify $\sigma(1, \boldsymbol{j})$ ($\boldsymbol{j} > 1$)

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

- From (10), we can identify $\sigma^2_\omega, \sigma^2_{\varepsilon(1)}, \sigma^2_\Upsilon$ up to $\sigma^2(1,1)$ for $\underline{\boldsymbol{a}}(1) \leq \boldsymbol{a} \leq \bar{\boldsymbol{a}}(1)$.
- Following Carneiro et al. (2003) and Heckman and Vytlacil (2007b), from the joint probabilities $\Pr(\boldsymbol{D}(1,\boldsymbol{a}) = \boldsymbol{d}(\boldsymbol{a}))$ and $\Pr(\boldsymbol{D}(1,\boldsymbol{a}') = \boldsymbol{d}(\boldsymbol{a}'))$, $\boldsymbol{d}(\boldsymbol{a}), \boldsymbol{d}(\boldsymbol{a}') \in \{0,1\}$, we can identify

$$\mathrm{Cov}\left(\frac{\Psi(1,\boldsymbol{a})}{\sigma(1,\boldsymbol{a})}, \frac{\Psi(1,\boldsymbol{a}')}{\sigma(1,\boldsymbol{a}')}\right) = \frac{(\boldsymbol{a} - \underline{\boldsymbol{a}}(1))(\boldsymbol{a}' - \underline{\boldsymbol{a}}(1))\sigma^2_\omega}{\sigma(1,\boldsymbol{a})\sigma(1,\boldsymbol{a}')} + \frac{\min((\boldsymbol{a} - \underline{\boldsymbol{a}}(1)), (\boldsymbol{a}' - \underline{\boldsymbol{a}}(1))\sigma^2_{\varepsilon(1)} + \sigma^2_\Upsilon}{\sigma(1,\boldsymbol{a})\sigma(1,\boldsymbol{a}')}. \tag{10}$$

- See Heckman and Vytlacil (2007a).

H.4 Identification of Higher Level Parameters

- The same logic extends to higher levels of ℓ, $\boldsymbol{L} \geq \ell > 1$, except here it is fruitful to distinguish two cases:
  1. With skill invariance
  2. Without skill invariance
- We first assume skill invariance over levels.
- We maintain throughout skill invariance within the same level.

With Skill Invariance

- We can write $\boldsymbol{K}(\ell, \boldsymbol{a})$ as

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

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

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

with covariance

$$\begin{aligned} \mathrm{Cov}(\Psi_\ell(\boldsymbol{a}), \Psi_\ell(\boldsymbol{a}')) &= (\boldsymbol{a} - \underline{\boldsymbol{a}}(1))(\boldsymbol{a}' - \underline{\boldsymbol{a}}(1))\sigma^2_\omega && (11) \\ &+ \sum_{k=1}^{\ell-1} \sigma^2_{\varepsilon(k)}(\bar{\boldsymbol{a}}(k) - \underline{\boldsymbol{a}}(k)) && (12) \\ &+ \min(\boldsymbol{a} - \underline{\boldsymbol{a}}(\ell), \boldsymbol{a}' - \underline{\boldsymbol{a}}(\ell))\sigma^2_{\varepsilon(\ell)} + \sigma^2_\Upsilon. && (13) \end{aligned}$$

- From each indicator variable, we can identify for each ℓ and $\boldsymbol{\alpha}$, the threshold variables $\frac{\bar{K}_\ell}{\sigma(\ell,\boldsymbol{\alpha})}$ and $\frac{\mu_0(\boldsymbol{Z})}{\sigma(\ell,\boldsymbol{\alpha})}$, $\frac{\bar{\beta}_\ell(\boldsymbol{X})}{\sigma(\ell,\boldsymbol{\alpha})}$, $\frac{\Delta_\ell(\boldsymbol{\alpha})}{\sigma(\ell,\boldsymbol{\alpha})}$, $\frac{\Lambda_\ell(\boldsymbol{\alpha})}{\sigma(\ell,\boldsymbol{\alpha})}$.
- The other terms in the index $\boldsymbol{K}(\ell, \boldsymbol{\alpha})$ are identified by a recursive argument starting from $\ell = 1$ (previously discussed).
- In Equation (11), the only unknown parameter is $\sigma^2_{\varepsilon(\ell)}$ and given the covarince, we can identify the variance term $\sigma^2_{\varepsilon(\ell)}$.
- Therefore, we can identify the variance of the sum of shocks $\Psi_\ell(\boldsymbol{\alpha})$ without imposing additional normalization.

Without Skill Invariance

- Suppose that we assume an affine transformation of the latent skill
$$\ln \boldsymbol{K}(\ell, \underline{\boldsymbol{a}}(\ell)) = \gamma_{0,\ell} + \gamma_{1,\ell} \ln \boldsymbol{K}(\ell - 1, \bar{\boldsymbol{a}}(\ell - 1))$$
- It is clearly impossible to separate $\gamma_{0,\ell}$ from the threshold parameter $\ln \bar{\boldsymbol{K}}(\ell)$ so we normalize $\gamma_{0,\ell} = 0$ with the understanding that any estimated threshold parameter is net of $\gamma_{0,\ell}$ (i.e., $\bar{\boldsymbol{K}}_\ell - \gamma_{0,\ell}$).
- To identify the parameter $\gamma_{1,\ell}$, we need to use the additional moment for the tasks across difficulty levels.

- In Equation (14) we have $\mathrm{Cov}\big(\Psi_\ell(\boldsymbol{a}), \Psi_{\ell'}(\boldsymbol{a}')\big)$ as following:

$$\mathrm{Cov}\big(\Psi_\ell(\boldsymbol{a}), \Psi_{\ell'}(\boldsymbol{a}')\big) = \gamma_{1,\ell}\Bigg\{(\boldsymbol{a} - \underline{\boldsymbol{a}}(1))(\boldsymbol{a}' - \underline{\boldsymbol{a}}(1))\sigma^2_\omega + \sigma^2_\Upsilon \quad (14)$$

$$+ \sum_{\boldsymbol{k}=1}^{\ell'-2}(\bar{\boldsymbol{a}}(\boldsymbol{k}) - \underline{\boldsymbol{a}}(\boldsymbol{k}))\sigma^2_{\varepsilon(\boldsymbol{k})}$$

$$+ (\boldsymbol{a}' - \underline{\boldsymbol{a}}(\ell' - \boldsymbol{a}))\sigma^2_{\varepsilon(\ell-1)}\Bigg\}.$$

- We can identify $\gamma_{1,\ell}$ based on this covariance term between level ℓ and level $\ell - 1$.
- Given similar logic, we can identify all parameters $\gamma_{1,\ell}$ across difficulty levels.
- This version of age invariance can be tested.

Beyond Normality

- Drawing on the analysis of Heckman and Vytlacil (2007a) and Matzkin (2007), with sufficient variation in the regressors, we can not only identify the normal distribution, but we can also identify the model under more general distributional assumptions based on Matzkin(1993).
- To identify all the shock distributions, we need to discuss how to identify the term $\bar{\Psi}_1(\boldsymbol{\alpha})$ first. Therefore, we need to impose some assumptions on $\mu_1 + \mu_0(\boldsymbol{Z}) + \boldsymbol{V}(\boldsymbol{Q}(\boldsymbol{\alpha})) + \bar{\beta}_1(\boldsymbol{X})\boldsymbol{\alpha}$ and the shock terms. We need to impose the following conditions
    - (a) this function is continuous function;
    - (b) there exists some $\boldsymbol{x}'$ and $\boldsymbol{z}'$ such that $\mu_1 + \mu_0(\boldsymbol{z}') + \boldsymbol{V}(\boldsymbol{Q}(\boldsymbol{\alpha})) + \bar{\beta}_1(\boldsymbol{x}')\boldsymbol{\alpha} = \boldsymbol{R}$ regardless the estimated coefficients
    - (c) the independence assumption between the shock term $\bar{\Psi}_1(\boldsymbol{\alpha})$ and the observable covariates.
    - (d) the distributions of all shocks are continuous.

All above assumptions (a-d) guarantee to nonparametrically identify $\bar{\Psi}_1(a)$. Next, we need to discuss how to separately identify the distribution of ω, ε and Υ. We can use characteristic function to discuss the identification.

- Denote the characteristic function of $\bar{\Psi}_1(a)$ is $\varphi_{\Psi(a)}(t)$
- Based on the assumption that all the shock terms are independent with each other we have the following: $\varphi_{\Psi(a)}(t) = \varphi_\omega(at)\varphi_{U_1(a)}(t)\varphi_\Upsilon(t)$
- Notice that $U_1(a) = \varepsilon(1,1) + \cdots + \varepsilon(1,a)$ Therefore, $\varphi_{U_1(a)}(t) = \varphi_{\varepsilon(1)}(t)^a$. So we can get the following equation:

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

If we impose ω distribution is standard normal distribution, we can identify

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

Therefore, characteristic function for $\varepsilon(1)$ can be identified by

$$\varphi_{\varepsilon(1)}(\boldsymbol{t}) = \frac{\varphi_{\Psi(1)}(\boldsymbol{t})}{\varphi_{\omega}(\boldsymbol{t})\varphi_{\Upsilon}(\boldsymbol{t})}$$

Since we can identify the shock distributions, we can easily get the first two orders of the moments. The identification of the remaining parameters follows from repeated application of the same logic.

Return to main text

Identification for Difficulty Level $\ell = 2$

- First we assume latent skill is invariant between level one and level 2. (i.e., $\ln \boldsymbol{K}(1, \bar{\boldsymbol{a}}(1)) = \ln \boldsymbol{K}(2, \underline{\boldsymbol{a}}(2))$)
- This assumption is equivalent to $\gamma_{0,2} = 0$, and $\gamma_{1,2} = 1$ for affine transformation. (We will relax this assumption later)
- The latent skill at level 2 at age $\boldsymbol{a}$ can be written as:

$$\ln \boldsymbol{K}(2, \boldsymbol{a}) = \boldsymbol{V}_2(\boldsymbol{a}) + \bar{\beta}_2(\boldsymbol{X})(\boldsymbol{a} - \bar{\boldsymbol{a}}(1)) + \sum_{j=\underline{\boldsymbol{a}}(2)}^{\boldsymbol{a}} \varepsilon(2, \boldsymbol{j}) + \ln \boldsymbol{K}(2, \underline{\boldsymbol{a}}(2))$$

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

Identification for Difficulty Level $\ell = 2$

- Similarly $\bar{\boldsymbol{K}}_2$, $\boldsymbol{V}_2(\boldsymbol{a})$ and $\bar{\beta}_2(\boldsymbol{X})$ can be identified up to scale $\boldsymbol{Var}(\Psi_2(\boldsymbol{a})) \equiv \sigma^2(2, \boldsymbol{a})$

$$\begin{aligned} \text{Cov}\left(\Psi_2(\boldsymbol{a}), \Psi_2(\boldsymbol{a}')\right) \quad &= \quad \sigma^2_{\Upsilon} + \boldsymbol{a}\boldsymbol{a}'\sigma^2_{\omega} + (\bar{\boldsymbol{a}}(1) - \underline{\boldsymbol{a}}(1))\sigma^2_{\varepsilon(1)} \\ &\quad +\min((\boldsymbol{a} - \underline{\boldsymbol{a}}(2)), (\boldsymbol{a}' - \underline{\boldsymbol{a}}(2)))\sigma^2_{\varepsilon(2)}. \end{aligned} \tag{15}$$

- Can identify $\sigma^2_{\varepsilon(2)}$ which is the only unknown parameters in the above equation, all of the other terms are identified at level 1.
- After identifying $\sigma^2_{\varepsilon(2)}$, can identify $\sigma(2, \boldsymbol{a})$.
- We do not need to impose an extra normalization at level 2 and can identify all parameters at level 2.

Return to main text

I. Affine Transformation

Supermodular Function If f is twice continuously differentiable, then f has increasing differences in (x, y) if and only if $y' \geq y$ implies that $f_x(x, y') \geq f_x(x, y)$ for all x, or alternatively that $f_{xy}(x, y) \geq 0$ for all x, y.

Now we want to show for twice continuous monotone transformation function $\phi(x, y)$ of function f: $R^k \mapsto R$, the affine transformation is the only monotone transformation.

- Let $\phi(x, y) = g \circ f(x, y)$,where $g(\cdot)$ is monotone and twice continuous function.
- If $\phi(x, y)$ is supermodular function, it should satisfy $\phi_{xy}(x, y) \geq 0$

$$\phi_{xy}(x, y) = g_{zz} f_x f_y + g_z f_{xy}$$

- If without additional assumption on $f_x(x, y)$ and $f_y(x, y)$, the only condition for $\phi(x, y)$ is supermodular function is $g_{zz} = 0$ and $g_z \geq 0$.
- Therefore only affine transformation satisfy this condition.

Return to main text

J. Simulation Procedure for Method of Moments Estimation

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

2. After drawing initial latent skills $\boldsymbol{K}(\boldsymbol{s}, 0)$, child $\boldsymbol{i}$ skills evolve following the equation

$$\ln \boldsymbol{K}_i(\boldsymbol{s}, \ell, \boldsymbol{a}) = \ln \boldsymbol{K}_i(\boldsymbol{s}, \ell, \boldsymbol{a}-1) + (\delta(\boldsymbol{s}, \ell)\eta_i(\boldsymbol{s}) + \boldsymbol{V}(\boldsymbol{Q}(\boldsymbol{a})) + \varepsilon_i(\boldsymbol{s}, \ell, \boldsymbol{a})),$$

where:

- $\delta(\boldsymbol{s}, \ell)$ is a level-specific lesson parameter for skill $\boldsymbol{s}$.
- $\varepsilon_i(\boldsymbol{s}, \ell, \boldsymbol{a})$ is an i.i.d. random term associated with the learning process drawn from a normal distribution $\boldsymbol{N}(0, \sigma^2(\boldsymbol{s}, \ell))$ for skill $\boldsymbol{s}$.
- $\eta_i(\boldsymbol{s})$ is an individual learning parameter with the specification. $\eta_i(\boldsymbol{s}) = \boldsymbol{X}\beta_{\boldsymbol{s}} + \gamma_i$, where $\boldsymbol{X}$ includes the set of interaction measures.
- γ_i is individual-specific learning ability, which follows the normal distribution $\boldsymbol{N}(0, \sigma^2_{\boldsymbol{s},\gamma})$.

3. Since $\delta(\mathsf{s}, \ell)$ and $\eta_i(\boldsymbol{x}(\boldsymbol{a}))$ enter the model multiplicatively, we cannot identify them separately without normalization.
   - We normalize the first level $\delta(\mathsf{s}, 1) = 1$ to identify the coefficients of β_{s}.
   - Since we assume that the β_{s} are the same across levels, we can identify $\delta(\mathsf{s}, \ell)$ for $\ell > 1$.
   - $\delta(\mathsf{s}, \ell)\eta_i(\boldsymbol{x}(\boldsymbol{a})) = \delta(\mathsf{s}, \ell)(\boldsymbol{X}'(\boldsymbol{a})\beta_{\mathsf{s}} + \gamma_i)$, where $\boldsymbol{X}$ includes determinants of ability and $\delta(\mathsf{s}, \ell)$ captures lesson content.

4. We now give an example of how to simulate the latent skills within the same difficulty level ℓ:
   1. Use the initial latent skill $\ln \boldsymbol{K}_i(\boldsymbol{s}, 0)$, which is formed by $\boldsymbol{Z}_i'\beta_{0,\boldsymbol{s}} + \tau_{i,0,\boldsymbol{s}}$ and the random draw $\tau_{i,0,\boldsymbol{s}}$ from a normal distribution with mean zero and variance $\sigma_\tau^2(0, \boldsymbol{s})$. Child ages sometimes differ at enrollment.
   2. If child $\boldsymbol{i}$ starts the intervention at difficulty level ℓ, we randomly draw the task error term at difficulty level ℓ from the normal distribution with mean zero and variance $\sigma^2(\varepsilon, \ell)$. Then, we can construct the latent skill for the first task based on the following equation:

$$\ln \boldsymbol{K}_i(\boldsymbol{s}, \ell, 1) = \ln \boldsymbol{K}(\boldsymbol{s}, 0) + \delta(\boldsymbol{s}, \ell)(\boldsymbol{X}'(1)\beta_{\boldsymbol{s}} + \gamma_i) + \varepsilon_i(\varepsilon, \ell, 1).$$

3. Similarly, we randomly draw a shock $\varepsilon_i(s, \ell, 2)$ for the 2nd task at level ℓ, and then construct the latent skill as follows:

$$\ln K_i(s, \ell, 2) = \ln K_i(s, \ell, 1) + \delta(s, \ell)(X'(2)\beta_s + \gamma_i) + \varepsilon_i(s, \ell, 2).$$

4. Repeat the previous step until the last task of difficulty level ℓ at age $\bar{a}(s, \ell)$ to construct the latent skill:

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

5. For a model with scale invariance, the latent skill across difficulty levels evolves as follows:

$$\begin{aligned}\ln \boldsymbol{K}_i(\mathbf{s}, \ell+1, \underline{\boldsymbol{a}}(\mathbf{s}, \ell+1)) = \ln \boldsymbol{K}_i(\mathbf{s}, \ell, \bar{\boldsymbol{a}}(\mathbf{s}, \ell)) \\ + \delta(\mathbf{s}, \ell+1)(\boldsymbol{X}'(\boldsymbol{a})\beta_{\mathbf{s}} + \gamma_i) \\ + \varepsilon_i(\mathbf{s}, \ell+1, \boldsymbol{a}(\mathbf{s}, \ell+1)).\end{aligned}$$

After the first task at that level, the latent skill evolves as follows:

$$\begin{aligned}\ln \boldsymbol{K}_i(\mathbf{s}, \ell+1, \boldsymbol{a}) = \ln \boldsymbol{K}_i(\mathbf{s}, \ell+1, \boldsymbol{a}-1) \\ + \delta(\mathbf{s}, \ell+1)(\boldsymbol{X}'(\boldsymbol{a})\beta_{\mathbf{s}} + \gamma_i) \\ + \varepsilon_i(\mathbf{s}, \ell+1, \boldsymbol{j}), \quad \boldsymbol{j} > 1.\end{aligned}$$

6. For a model Without Age Invariance, the latent skill across difficulty levels develops as follows:

$$\begin{aligned}\ln \boldsymbol{K}_i(\mathbf{s}, \ell+1, \boldsymbol{a}(\mathbf{s}, \ell, 1)) = {} & \gamma_{0,\ell} + \gamma_{1,\ell}(\ln \boldsymbol{K}_i(\mathbf{s}, \ell)) \\ & + \delta(\mathbf{s}, \ell+1)(\boldsymbol{X}'(1)\beta_s + \gamma_i) \\ & + \varepsilon_i(\mathbf{s}, \ell+1, \boldsymbol{a}(\mathbf{s}, \ell+1)).\end{aligned}$$

After the first task, the latent skill at other tasks at level $\ell+1$ is as follows:

$$\begin{aligned}\ln \boldsymbol{K}_i(\mathbf{s}, \ell+1, \boldsymbol{a}(\mathbf{s}, \ell+1, \boldsymbol{j})) = {} & \ln \boldsymbol{K}_i(\mathbf{s}, \ell+1, \boldsymbol{a}(\mathbf{s}, \ell+1, \boldsymbol{j}-1)) \\ & + \delta(\mathbf{s}, \ell+1)(\boldsymbol{X}'(\boldsymbol{j})\beta_s + \gamma_i) \\ & + \varepsilon_i(\mathbf{s}, \ell+1, \boldsymbol{j})), \quad \boldsymbol{j} > 1.\end{aligned}$$

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

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

8. We form moments based on a series of simulated child task performance and minimize the distance between the simulated and empirical moments.

9. The moments we consider in estimation include:
   1. All task passing rates;
   2. The passing rate on the first five tasks at each level;
   3. The passing rate for each difficulty level;
   4. The passing rate for newly enrolled children;
   5. The passing rate for those who enroll in the program longer than one month;

6. The probability of passing the j'th task ($j \neq j'$) within each level, conditioning on the child passing the jth task;
7. The probability of passing the j'th task at level $\ell + 1$ across all difficulty levels, conditioning on the child passing the jth task at level ℓ;
8. The probability of passing the j'th task at level $\ell + 2$, conditioning on the child passing the jth task at level ℓ.

10. After obtaining the point estimates, we bootstrap to calculate the standard errors of the estimates.

Table 44: Assumptions on Random Shocks in the Model

Parameters	Level Specific or Not	Distribution
Initial Latent Skill Condition Shock ($\tau_i(0, \mathbf{s})$)	No	Normal$(0, \sigma^2_\tau(0, \mathbf{s}))$
Learning ability Shock ($\omega_i(\mathbf{s})$)	No	Normal$(0, \sigma^2_\omega(0, \mathbf{s}))$
Task Performance Shock ($\varepsilon_i(\mathbf{s}, \ell, \boldsymbol{\alpha})$)	Yes	Normal$(0, \sigma^2(\varepsilon, \mathbf{s}, \ell))$

J.1 Bootstrap Procedure

- Since our data are clustered at the village level, to conduct robust inference, we use the paired cluster bootstrap method in our paper.
- The paired cluster bootstrap procedure is as follows:
  1. From the original sample, we get point estimates β^*.
  2. We iterate the following bootstrap procedure 1000 times from a sample of G clusters $(\boldsymbol{y}_1, \mathbf{X}_1), \cdots, (\boldsymbol{y}_G, \mathbf{X}_G)$, resampling with replacement G times from the original sample of clusters. The unit of bootstrap is at cluster level. After we randomly draw $\boldsymbol{G}$ clusters, we construct one bootstrap sample.

3. Based on the bootstrap sample $\boldsymbol{k}$, we estimate the structural model based on the estimation procedure documented above and get the point estimates $\beta_{\boldsymbol{k}}^{\boldsymbol{bs}}$ for the structural model.
4. We conduct inference on each estimated parameter β^* based on the distribution of $\beta^{\boldsymbol{bs}}$. After iterating the bootstrap 1000 times, we have the distribution of each parameter. We then calculate the confidence interval and standard error for each parameter.

Return to main text

K. Estimation: Moment Fit

Table 45: Moments Used in Estimation

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

- In summary, 80% of the simulated moments are in the 95% confidence intervals of data moments.
- Overall, our estimates fit the moments very well.
- The model Without Age Invariance has better fit.
- We also examine the model of fit by the following summary measure:

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

where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment i.

Table 46: Goodness of Fit Summary

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

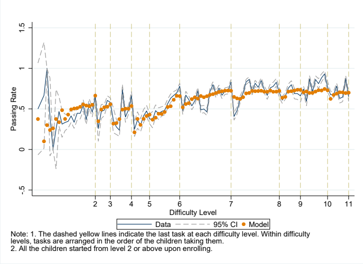
1. $R = \frac{\sum_i (y_i^m - y_i^d)^2}{N_I}$, where y_i^m is the predicted moment i for the model, and y_i^d is the empirical moment.

2. We cannot reject the model at the 0.0001 level.

K.1 Language

Figure 33: Fit for All Language Tasks

(a) With Scale Invariance



(b) Without Age Invariance

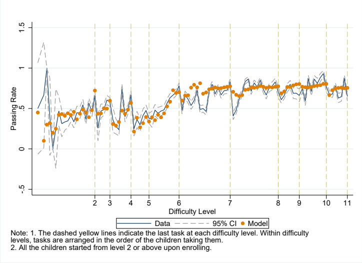
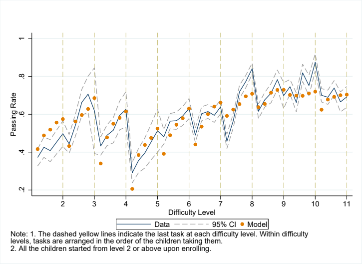


Figure 34: Fit for the First Five Tasks at Each Level

(a) With Scale Invariance



(b) Without Age Invariance

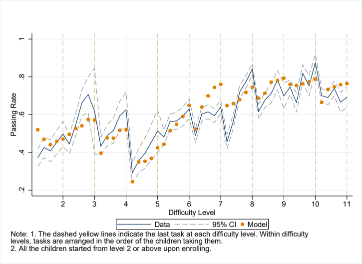
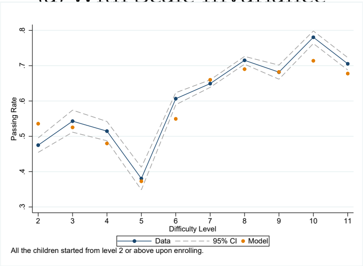


Figure 35: Fit by Level

(a) With Scale Invariance



(b) Without Age Invariance

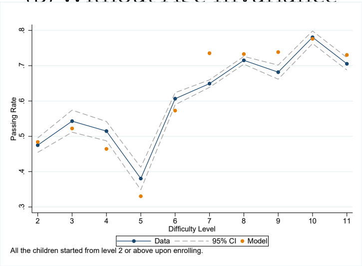
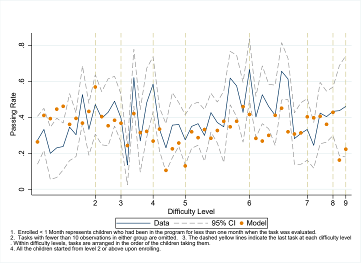


Figure 36: Fit by Length of Enrollment: Newly-Enrolled Group

(a) With Age Invariance



(b) Without Age Invariance

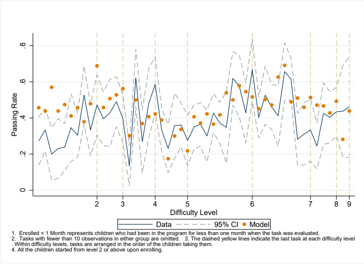
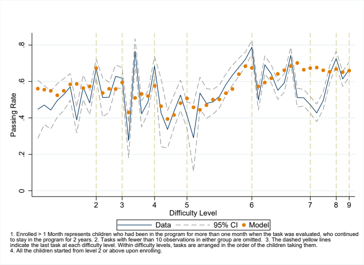


Figure 37: Fit by Length of Enrollment: Group Enrolled > 1 Month

(a) With Age Invariance



(b) Without Age Invariance

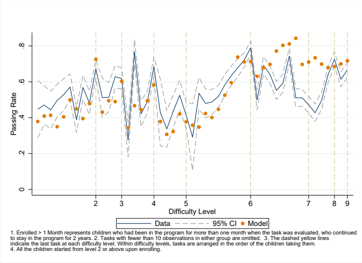
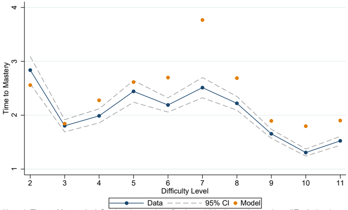


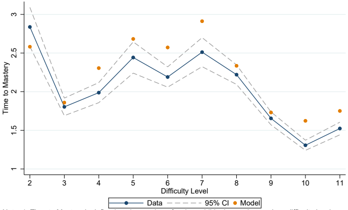
Figure 38: Fit for Time to Mastery

(a) With Age Invariance



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

(b) Without Age Invariance

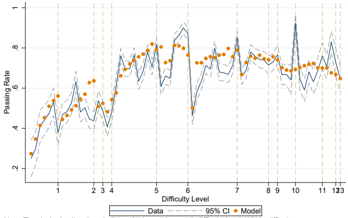


Note: 1. Time to Mastery is defined as the number of tasks a child takes at the previous difficulty level until the first success (inclusive).
2. All the children started from level 2 or above upon enrolling.

K.2 Cognition

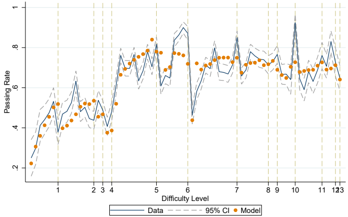
Figure 39: Fit for All Cognitive Tasks

(a) With Age Invariance



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

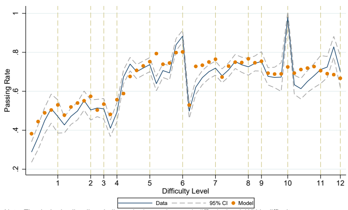
(b) Without Age Invariance



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

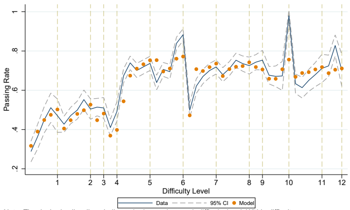
Figure 40: Fit for the First Five Tasks at Each Level

(a) With Age Invariance



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

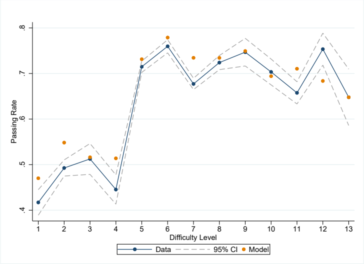
(b) Without Age Invariance



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

Figure 41: Fit by Level

(a) With Age Invariance



(b) Without Age Invariance

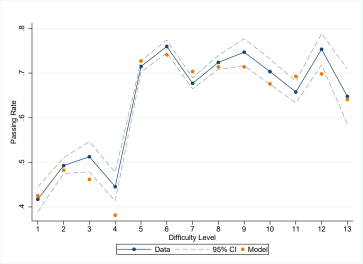
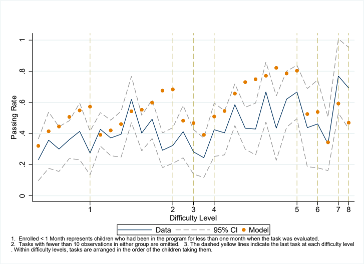


Figure 42: Fit by Length of Enrollment: Newly-Enrolled Group

(a) With Age Invariance



(b) Without Age Invariance

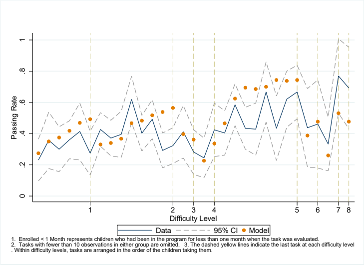
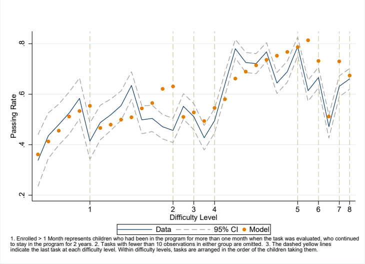


Figure 43: Fit by Length of Enrollment: Group Enrolled > 1 Month

(a) With Age Invariance



(b) Without Age Invariance

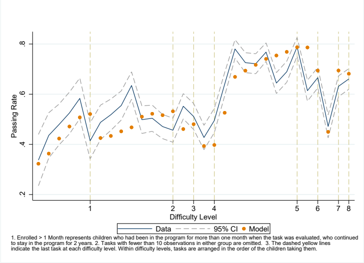
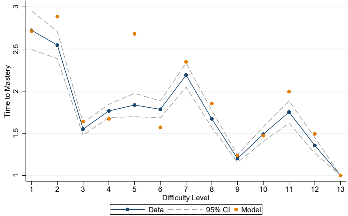


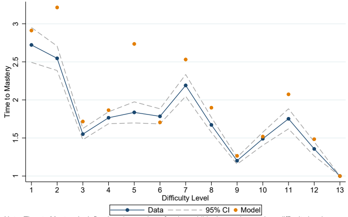
Figure 44: Fit for Time to Mastery

(a) With Age Invariance



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

(b) Without Age Invariance

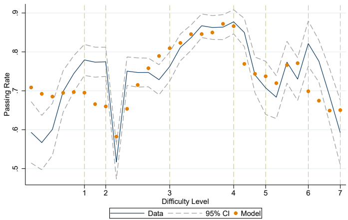


Note: Time to Mastery is defined as the number of tasks a child takes at the previous difficulty level until the first success (inclusive).

K.3 Fine Motor

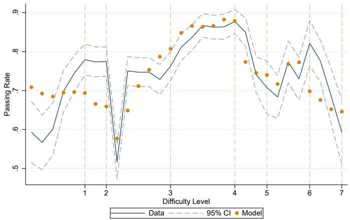
Figure 45: Fit for All Fine Motor Tasks

(a) With Age Invariance



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

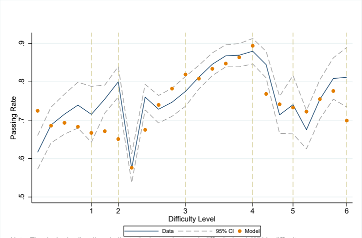
(b) Without Age Invariance



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

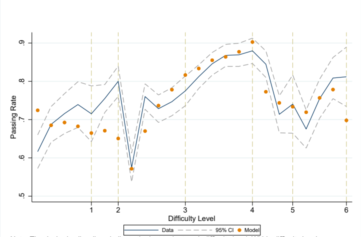
Figure 46: Fit for the First Five Tasks at Each Level

(a) With Age Invariance



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

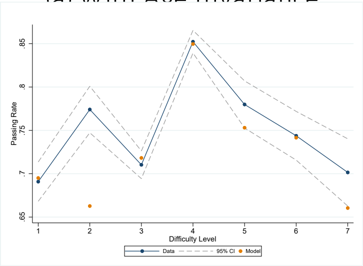
(b) Without Age Invariance



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

Figure 47: Fit by Level

(a) With Age Invariance



(b) Without Age Invariance

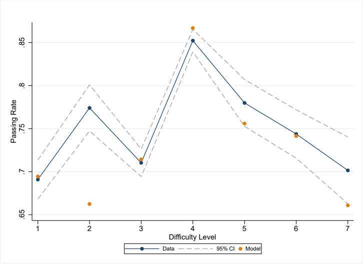
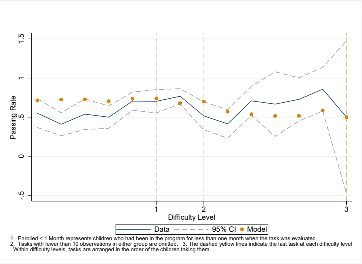


Figure 48: Fit by Length of Enrollment: Newly-Enrolled Group

(a) With Age Invariance



(b) Without Age Invariance

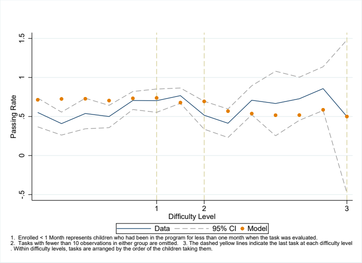
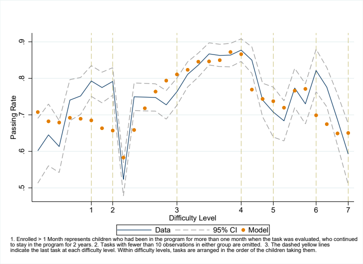


Figure 49: Fit by Length of Enrollment: Group Enrolled > 1 Month

(a) With Age Invariance



(b) Without Age Invariance

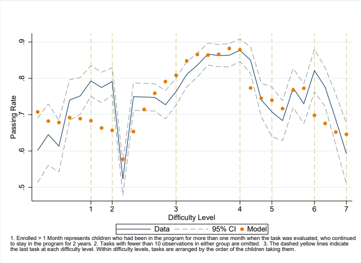
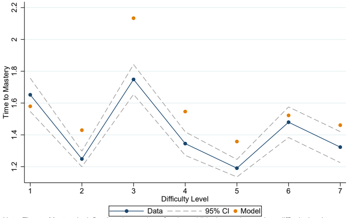


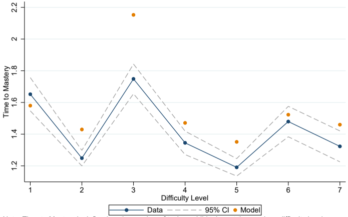
Figure 50: Fit for Time to Mastery

(a) With Age Invariance



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

(b) Without Age Invariance



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

K.4 Language Skill Moment Fit Summary

- Overall, our estimates fit the moments very well.
- The model Without Age Invariance has better fit.
- We also examine the model of fit by the following summary measure:

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

where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment i.

Table 47: Goodness of Fit Summary (Language)

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

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

2. We cannot reject the model at the 0.0001 level.

K.5 Cognitive Skill Moment Fit Summary

- Overall, our estimates fit the moments very well.
- The model Without Age Invariance has better fit.
- We also examine the model of fit by the following summary measure:

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

where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment i.

Table 48: Goodness of Fit Summary (Cognitive)

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

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

2. We cannot reject the model at the 0.0001 level.

K.6 Fine Motor Skill Moment Fit Summary

- Overall, our estimates fit the moments very well.
- The model Without Age Invariance has better fit.
- We also examine the model of fit by the following summary measures:

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

where y_i^m is a predicted moment i for the model, and y_i^d is the empirical moment i.

Table 49: Goodness of Fit Summary (Fine Motor)

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

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

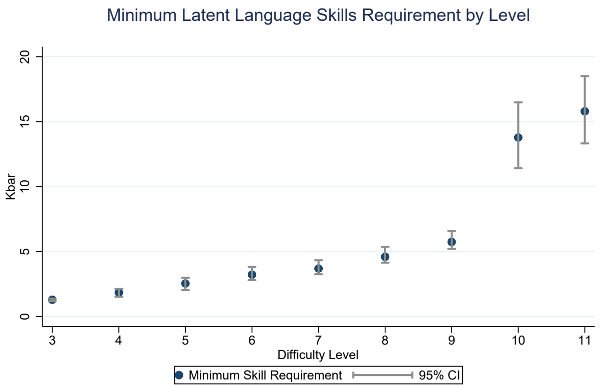
2. We cannot reject the model at the 0.0001 level.

Return to main text

L. Point Estimates

Language Skill

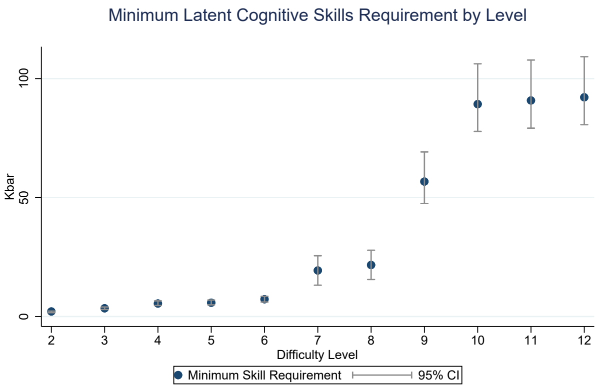
Figure 51: $\bar{K}(\ell)$



Note: The confidence interval is based on 1,000 iteration bootstrap.
All the children started from level 2 or above upon enrolling.
The value at Level 2 is normalized to one

Cognitive Skill

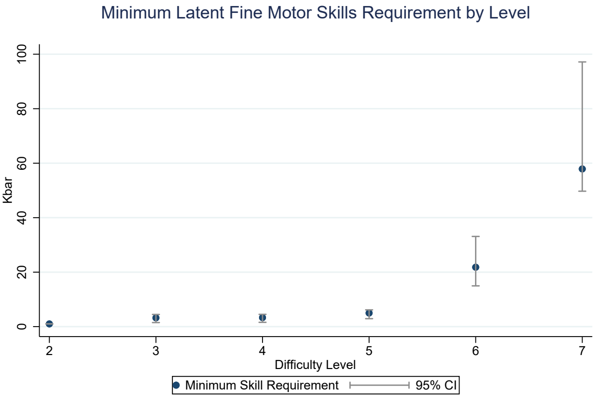
Figure 52: $\bar{K}(\ell)$



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

Fine Motor Skill

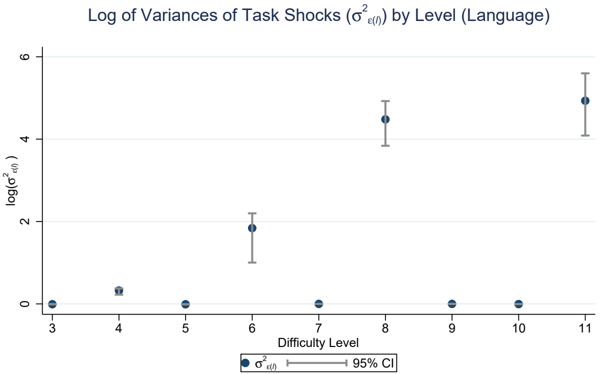
Figure 53: $\bar{K}(\ell)$



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

Language Skill

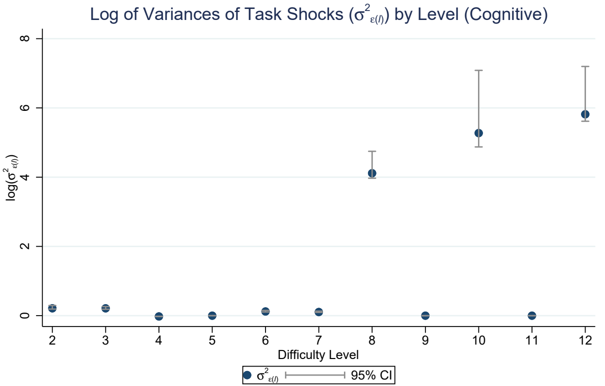
Figure 54: $\sigma^2_{\varepsilon}(\ell)$



Note: The confidence interval is based on 1,000 iteration bootstrap.
Log values are shown in figures to better visualize the 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

Cognitive Skill

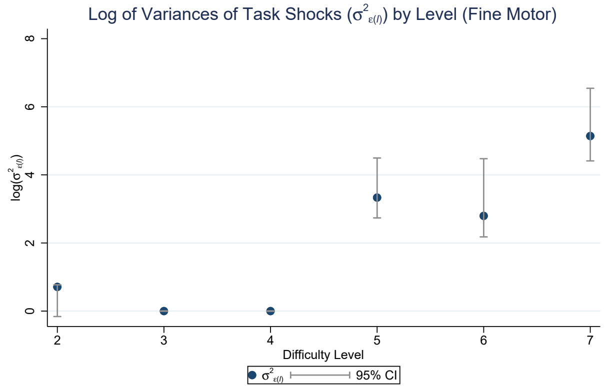
Figure 55: $\sigma_{\varepsilon}^{2}(\ell)$



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

Fine Motor Skill

Figure 56: $\sigma^2_\varepsilon(\ell)$



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

Return to main text

Table 50: Skill Invariance Hypothesis Tests by Levels

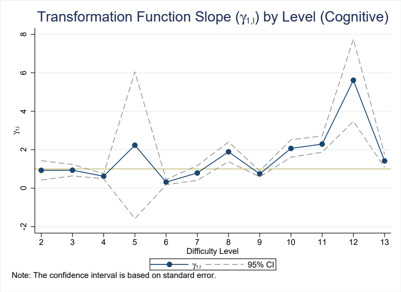
	Language			Cognitive			Fine Motor		
	Slope($\gamma_{1,\ell}$)	$\chi^2(\cdot)$	p-value	Slope($\gamma_{1,\ell}$)	$\chi^2(\cdot)$	p-value	Slope($\gamma_{1,\ell}$)	$\chi^2(\cdot)$	p-value
Level 2				0.929	0.012	0.914	1.005	0.000	0.992
Level 3	0.901	0.546	0.460	0.936	0.010	0.922	0.963	0.022	0.883
Level 4	0.645	20.193	0.000	0.621	0.142	0.707	1.446	0.774	0.379
Level 5	0.66	9.382	0.002	2.235	3.899	0.048	0.798	0.720	0.396
Level 6	1.522	5.063	0.024	0.317	17.482	0.000	0.748	1.277	0.258
Level 7	1.125	0.182	0.670	0.791	0.362	0.547	0.955	0.034	0.853
Level 8	0.562	8.195	0.004	1.893	4.237	0.040			
Level 9	1.113	0.113	0.737	0.744	3.432	0.064			
Level 10	1.006	0.001	0.970	2.068	12.211	0.000			
Level 11	1.223	0.375	0.540	2.292	10.927	0.001			
Level 12				5.614	14.351	0.000			
Level 13				1.420	4.333	0.037			
Total		44.051	0.000		71.398	0.000		2.827	0.830

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

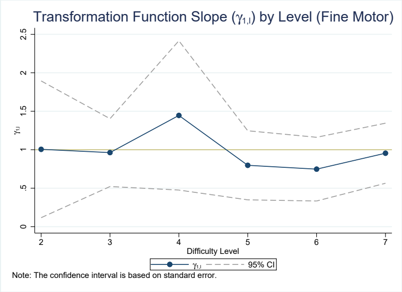
Testing Age Invariance

Figure 57: Tests of the Null Hypothesis of Age Invariance

Cognitive Skill



Fine Motor Skill



Return to main text

Table 51: Estimates of the Skill Transition Matrix

A_Lang-Lang	0.933***	A_Lang-Cog	0.002	A_Lang-Fine	0.015**
	(0.085)		(0.009)		(0.007)
A_Cog-Lang	0.050***	A_Cog-Cog	0.994***	A_Cog-Fine	0.038***
	(0.028)		(4.478)		(0.009)
A_Fine-Lang	-0.001	A_Fine-Cog	-0.001	A_Fine-Fine	1.028***
	(0.007)		(0.015)		(5.287)

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

- Cognitive skill will benefits both language and fine motor skill development in the next period.
- Cognitive and language skill will benefit fine motor skill development but fine motor skill itself cannot benefit language nor cognitive skill development

Return to main text