Measuring Knowledge

Including excerpts from "Early Childhood Learning Patterns for a Home Visiting Program in Rural China" by Jin Zhou, James Heckman, Fuyao Wang, and Bei Liu

James J. Heckman and Jin Zhou

Econ 350, Spring 2022

Introduction

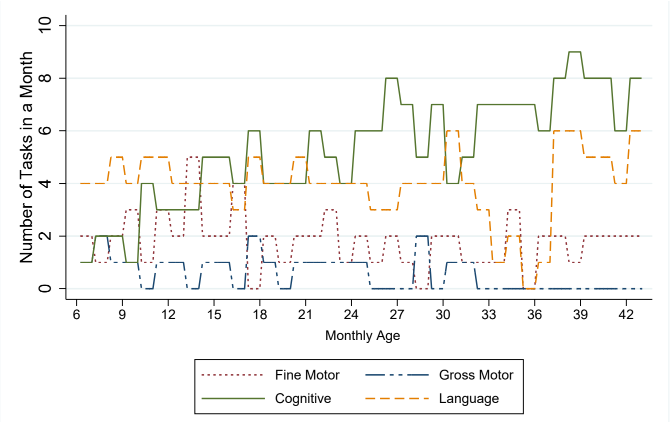
- A crucial assumption maintained in the literature on skill formation, ethnic skill gaps, and the economics of education is the existence of constant-unit latent skills ("human capital") over ages and inputs, which can be meaningfully compared across time and over people.
- A corollary but distinct assumption made in empirical work on measuring achievement growth and gaps and value-added measures is the existence of invariant measuring rods for latent skills, which may or may not exist even if there are true latent skill scales.
- This paper tests for the existence of such invariant measures for prototypical achievement and assessment tests.

- The central assumption in this paper is that mastery of tasks ***within a well-defined level*** is a true or foundational measure of knowledge.
- We can chart mastery within levels and compare knowledge and growth across children on a common microscale.
- Children can either perform a task successfully or not.
- We use this standard to assess the validity of more aggregative conventional measures of knowledge used in the economics of education and in the study of child development.
- Our study calls into question the conventional practice that relies on these aggregates as measures of knowledge that can be used to create meaningful comparisons across people or across time.

Our Measures of Skill

- ***China REACH*** was implemented in 2015 by a large-scale randomized control trial.
  - It enrolled 1,500 participants aged 9–30 months (about 700 participants in the treatment group) in 111 villages in Huachi county, Gansu province, one of the poorest areas of China.
  - Trained home visitors visit each treated household weekly and provide one hour of parenting or caregiving guidance.
  - Three or four different skills (gross motor, fine motor, language, and cognitive) are taught each week.
- We assume that ***knowledge content is the same within levels***.
- Figure 1 presents the skill tasks taught and measured at each age.

Figure 1: Curriculum Task Intensity: The Number of Tasks in a Month in the Curriculum (by Skill Category)



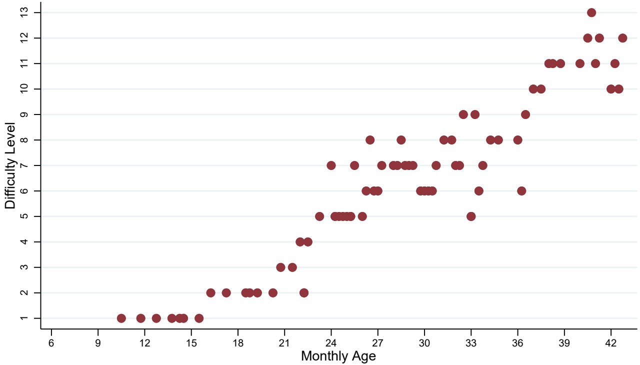
Cognitive Skills

- Cognitive skills have different dimensions.
- In the curriculum, the cognitive skills taught cover spatial skills, knowledge of objects and object functions, order and number, etc.
- We use knowledge of objects and object functions as an example.
- Cognitive skill difficulty levels are defined based on the abstract concepts shown in Table 1.
- Seventy-four lessons are sorted into the thirteen ordered difficulty levels.

Table 1: Difficulty Level List for the Cognitive Understanding Objects Lessons

Level 1	The child looks at the pictures and vocalizes.
Level 2	Name the objects and ask the child to point to the corresponding pictures.
Level 3	The child can point to one picture and name the objects in it.
Level 4	The child can point to two or more pictures and name the objects in them.
Level 5	The child can point to three or more pictures and name the objects in them.
Level 6	The child can point to six or more pictures and name the objects in them.
Level 7	The child can talk about the pictures, answer questions, and understand or name actions (eat, play, etc.).
Level 8	The child can follow the storyline, answer questions, and name actions.
Level 9	The child can understand stories and talk about the content of 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 graphic, discuss the role of each item, then link the graphics in the card together.
Level 12	The child can name the objects in the picture, link different pictures together, and discuss some of the activities in the pictures.
Level 13	The child can name the objects in the picture and talk about their functions.

Figure 2: The Timing of Cognitive Skill (Understanding Objects) Tasks across Difficulty Levels



Note: Level 1: Look at the pictures and vocalize; Level 13: The child can name the things in the picture and talk about the function of objects.

- Figure 2 shows the timing of the cognitive skill (knowing objects and understanding object functions) levels in the curriculum.
- The number of lessons varies across difficulty levels according to the curriculum content itself.
- As children age and advance across difficulty levels, they confront more demanding tasks.

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

Month	Week	Learning Materials	Content
10	2	Picture book A	The baby makes sounds when looking at the pictures.
11	3	Picture book B	The baby looks at the pictures and vocalizes.
12	3	Picture book A	The child makes sounds looking at the pictures.
13	3	Picture book B	The child makes sounds looking at the pictures.
14	1	Picture book A	Mother and child look at the pictures together, and the mother lets the child vocalize and touch the pictures.
15	2	Picture book B	Mother and child look at the pictures together, and the mother lets the child vocalize and touch the pictures.

- Table 2 presents detailed information about the six lessons (and assessments) that are labeled as difficulty level 1 directed to ten-month-old to fifteen-month-old children.
- All of the lessons relate to the activity of looking at the pictures or objects and vocalizing, which does not require the child to name or identify the object.

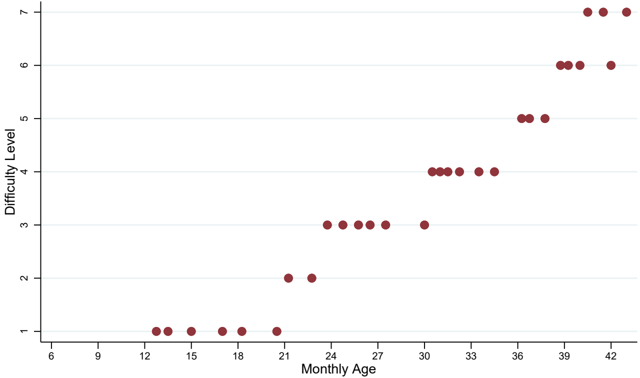
Fine Motor Skills

- Fine motor drawing lessons focus on a child's ability to use writing utensils on progressively more difficult tasks.
- First, a child is asked to hold utensils to make markings.
- The child is then asked to copy the markings made by an adult.
- As the skill levels progress, the child is asked to make markings after only hearing a verbal command from an adult.
- Finally, the child progresses from abstract shapes to representative drawing.
- See Table 3.

Table 3: Skill Levels for Fine Motor (Drawing) Lessons

Difficulty Level	Task Content
1	Doodle using crayons
2	Mimic circles
3	Mimic circles and draw straight lines
4	Draw a circle, vertical line, and horizontal line
5	Draw circles, many lines, and crossed lines
6	Draw a cross (or T), curves, and zigzag curves
7	Draw caterpillars

Figure 3: The Timing of Fine Motor Skill (Drawing) Tasks across Difficulty Levels



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

- Figure 3 gives the timing of each fine motor drawing assessment in the curriculum design.
- Difficulty level 1 covers the ages from 12 months and 3 weeks to 20 months and 2 weeks.
- In general, higher difficulty levels appear at later weekly ages.
- However, there can be some overlap across difficulty levels.
- When fine motor lessons at difficulty level 7 start, the student still receives lessons at difficulty level 6.
- Circling back is a strategy designed to solidify a child's understanding of a concept.

Our Key Identifying Assumption

- The curriculum we study targets lessons at different skill levels at each weekly age.
- For each type of skill, task difficulty levels are constructed following UHP.
- We use mastery of tasks at each level of skill as our fundamental measure of knowledge.
- Knowledge is acquired in real time.
- It may be forgotten or retained as children advance through the curriculum, leading to multiple measures of knowledge.
- Different types of knowledge can be acquired at different levels.

A Model for Measuring Knowledge

- Let $\mathcal{S}$ be the set of skills taught.
- Let $\ell(s, a) \in \{1, \ldots, L_s\}$ be the level of skill s taught at age a.
- L_s is the number of difficulty levels for each skill s.
- Mastery of skill s at level ℓ at age a is characterized by a threshold crossing model:

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

  where $D(s, \ell, a)$ records mastery (or not) of a skill s at a given level ℓ at age a, and $\bar{K}(s, \ell)$ is the minimum latent skill required to master the task at difficulty level ℓ.
- This characterization is consistent with the classical IRT model in educational psychology.

- Let $\underline{a}(s,\ell)$ be the first age at which skill s is measured at level ℓ, and $\bar{a}(s,\ell)$ be the last age at which it is measured at level ℓ.
- For consecutive lessons in a run, $1+\bar{a}(\ell)-\underline{a}(\ell)$ is the length of the run (# of lessons measured on skill s at level ℓ) starting at age $\underline{a}(s,\ell)$.
- For level ℓ of skill s, collect the indicators of knowledge in a spell:

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

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

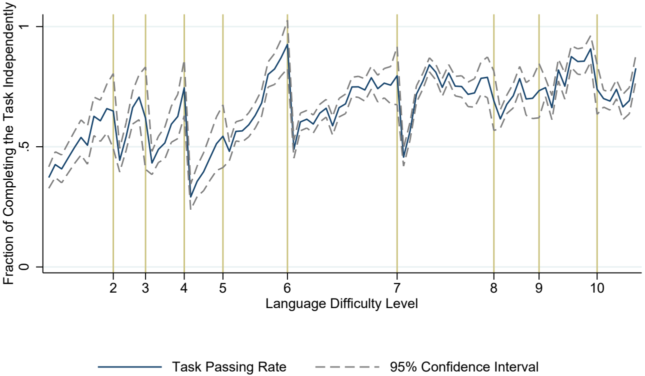
- With learning, sequences are back-loaded.
- For $j > 0$,
$$\Pr(D(s, \ell, a + j) \geq D(s, \ell, a)) \geq 0.$$
- Knowledge acquisition for each skill s at each level ℓ is measured by properties of these arrays and their relationships.
- Zhou, Heckman, Wang, and Liu (2021) test and reject the hypothesis of no learning for our data.
- They control for maturation and exposure effects that might boost skills in the absence of any intervention.
- Even after doing so, they reject exchangeability and find evidence of knowledge growth throughout the program.

THE UNIVERSITY OF CHICAGO

- Figure 4 characterizes the growth of knowledge of language, cognitive, and fine motor skills.
- Average passing rates within each difficulty level for language and cognitive tasks increase with age, a pattern consistent with learning.
- When individuals transition to a higher difficulty level, initial passing rates decline.
- Subsequent passing rates increase as learning ensues.

Figure 4: Average Task Passing Rate by Order and Level

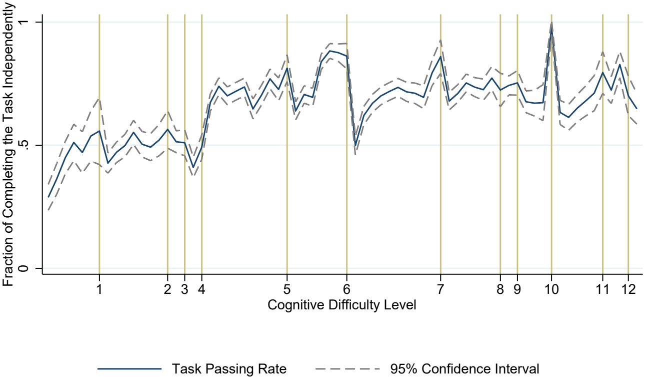
(a) Language*



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

Figure 4: Average Task Passing Rate by Order and Level, Cont'd

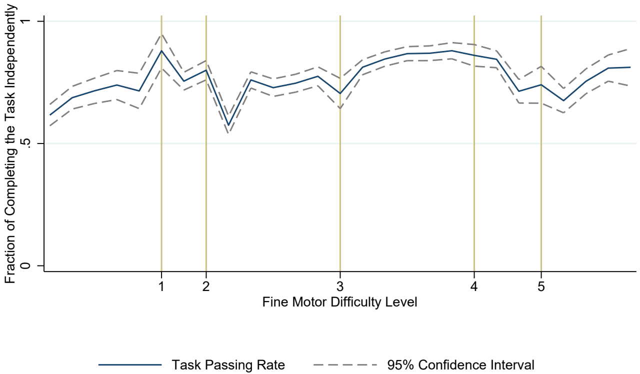
(b) Cognitive



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

Figure 4: Average Task Passing Rate by Order and Level, Cont'd

(c) Fine Motor



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

Measures of Knowledge and Knowledge Acquisition

- The traditional measure of knowledge of a skill is the proportion of correct answers over all levels of difficulty.
- A more refined measure within an assessment is defined within a skill and difficulty level (s, ℓ).
- The passing rate on skill s at level ℓ is:

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

- The overall passing rate is:

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

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

- This measure is an aggregate measure that does not recognize the sampling of (s, ℓ) items, the retention of knowledge, or the speed of acquisition.

- We define other plausible measures of knowledge and knowledge acquisition, which we also measure.
- For consecutive learning spells with all participants entering each level at the first lesson:
    - ***Time to first mastery*** is $d(s,\ell) = \hat{a}(s,\ell) - \underline{a}(s,\ell)$, where for each s and ℓ, $\hat{a}(s,\ell) = \min_a \{D(s,\ell,a) = 1\}_{a=\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}$.
    - ***Time to full mastery*** is $\tilde{a}(s,\ell) - \underline{a}(s,\ell)$.
- Some would call speed of mastery an ability and not a pure measure of knowledge.
- Other measures of learning are possible, such as time to mastery of two items in a row after $\hat{a}(s,\ell)$, etc.

- ***Backsliding*** at level ℓ for skill s is

$$\frac{\#\{D(s,\ell,a)=0, a>\hat{a}(s,\ell), a\leq\bar{a}(s,\ell)\}}{\#\{a>\hat{a}(s,\ell), a\leq\bar{a}(s,\ell)\}}\mathbf{1}(\#\{a>\hat{a}(s,\ell), a\leq\bar{a}(s,\ell)\}>0).$$

Link to "Early Childhood Learning Patterns" Extract

Correlations with Conventional Test Scores

- It is instructive to examine the correlation between the measures just defined and traditional achievement scores.
- We use Denver tests, which are closely related to the Bayley tests used to measure child development, as traditional scores.
- Tables 4a–4d present the correlations between the Denver scores at midline and endline and the average passing rate (the common measure of "knowledge") cumulated up to the date at which the Denver test is administered.

Table 4a: Correlation between Average Passing Rate (Up to Midline/Endline Measurement Age) and Denver Scores

		Average Passing Rate			
		Language	Cognitive	Fine Motor	Gross Motor
Denver Score (Midline)	Language and Cognitive	0.039**	0.078***	0.061**	0.043**
	Fine Motor	0.040**	0.076***	0.057**	0.086***
	Gross Motor	0.027	0.080***	0.054*	0.011
	Socioemotional	0.100***	0.118***	0.068**	0.068***
Denver Score (Endline)	Language and Cognitive	0.078***	0.098***	0.099***	0.058***
	Fine Motor	0.011	0.042***	0.042**	0.017
	Gross Motor	0.075***	0.088***	0.064***	0.055***
	Socioemotional	0.005	0.024*	0.044**	-0.001

Table 4b: Correlation between Time to First Mastery (Up to Midline/Endline Measurement Age) and Denver Scores

		Time to First Mastery			
		Language	Cognitive	Fine Motor	Gross Motor
Denver Score (Midline)	Language and Cognitive	-0.044**	-0.064***	-0.081***	-0.048**
	Fine Motor	-0.044**	-0.043**	-0.054*	-0.049**
	Gross Motor	-0.030	-0.078***	-0.034	-0.008
	Socioemotional	-0.071***	-0.073***	-0.060**	0.000
Denver Score (Endline)	Language and Cognitive	-0.076***	-0.069***	-0.052**	0.019
	Fine Motor	-0.024	-0.027*	-0.017	-0.002
	Gross Motor	-0.071***	-0.071***	-0.012	-0.027
	Socioemotional	-0.020	-0.023	0.029	0.003

Table 4c: Correlation between Instability (Up to Midline/Endline Measurement Age) and Denver Scores

		Instability			
		Language	Cognitive	Fine Motor	Gross Motor
Denver Score (Midline)	Language and Cognitive	-0.049**	-0.110***	-0.101***	-0.063**
	Fine Motor	-0.032	-0.058**	-0.058*	-0.103***
	Gross Motor	-0.023	-0.033	-0.101***	-0.032
	Socioemotional	-0.022	-0.094***	-0.050	-0.038
Denver Score (Endline)	Language and Cognitive	-0.070***	-0.063***	-0.043*	-0.078***
	Fine Motor	-0.026	-0.040**	-0.021	-0.031
	Gross Motor	-0.061***	-0.074***	-0.048**	-0.061**
	Socioemotional	0.003	-0.019	-0.041*	-0.032

Table 4d: Correlation between Time to Full Mastery (Up to Midline/Endline Measurement Age) and Denver Scores

		Time to Full Mastery			
		Language	Cognitive	Fine Motor	Gross Motor
Denver Score (Midline)	Language and Cognitive	-0.062***	-0.076***	-0.126***	-0.015
	Fine Motor	-0.040**	-0.034	-0.033	-0.035
	Gross Motor	-0.010	-0.025	-0.085**	0.031
	Socioemotional	-0.022	-0.029	-0.028	0.008
Denver Score (Endline)	Language and Cognitive	-0.049***	-0.046**	-0.082***	-0.078**
	Fine Motor	-0.022	-0.036**	-0.070**	-0.050
	Gross Motor	-0.030	-0.024	-0.020	-0.066**
	Socioemotional	-0.028	-0.001	-0.027	-0.044

- Most of the measures are significantly correlated with the children's Denver test scores in the expected directions.
- The Denver score is positively correlated with the average passing rate across tasks during the intervention.
- Notice, however, the strong correlations between Denver tasks tailored to a particular skill and the components of knowledge from all skills.
- This might suggest a one-dimensional model of skill.
- However, we test and reject that model.

- In addition to correlating knowledge measured over intervals, it is useful to measure knowledge at the time the Denver tests are taken.
- Tables 5a–5d report such correlations.
- The contemporaneous measures of knowledge are much more weakly correlated with the Denver scores.
- Cumulative measures are more predictive.

Table 5a: Correlation between Average Passing Rate (At Midline/Endline Measurement Age) and Denver Scores

		Average Passing Rate			
		Language	Cognitive	Fine Motor	Gross Motor
Denver Score (Midline)	Language and Cognitive	0.101**	0.074*	0.100	0.050
	Fine Motor	0.149***	0.069	0.170***	0.097*
	Gross Motor	0.147***	0.062	0.142**	0.012
	Socioemotional	0.128***	0.043	0.066	0.012
Denver Score (Endline)	Language and Cognitive	0.004	0.127*	0.058	-0.076
	Fine Motor	-0.249**	-0.066	-0.086	0.308
	Gross Motor	-0.085	0.198***	0.057	0.118
	Socioemotional	-0.216*	0.129**	0.115	0.078

Table 5b: Correlation between Time to First Mastery (At Midline/Endline Measurement Age) and Denver Scores

		Time to First Mastery			
		Language	Cognitive	Fine Motor	Gross Motor
Denver Score (Midline)	Language and Cognitive	-0.056	0.072	-0.045	-0.046
	Fine Motor	-0.052	0.012	0.006	0.018
	Gross Motor	-0.085*	0.013	-0.069	0.045
	Socioemotional	-0.039	-0.032	0.017	-0.013
Denver Score (Endline)	Language and Cognitive	0.091	-0.114	-0.004	0.076
	Fine Motor	-0.026	-0.010	0.038	-0.308
	Gross Motor	-0.049	-0.207***	0.047	-0.118
	Socioemotional	0.187	-0.250***	0.034	-0.078

THE UNIVERSITY OF CHICAGO

Table 5c: Correlation between Instability (At Midline/Endline Measurement Age) and Denver Scores

		Instability			
		Language	Cognitive	Fine Motor	Gross Motor
Denver Score (Midline)	Language and Cognitive	-0.148***	-0.074	-0.044	0.044
	Fine Motor	-0.049	-0.056	-0.091	-0.025
	Gross Motor	-0.004	-0.004	-0.019	0.048
	Socioemotional	-0.061	-0.026	0.012	0.129*
Denver Score (Endline)	Language and Cognitive	-0.294*	-0.025	0.064	.
	Fine Motor	0.069	0.086	0.026	.
	Gross Motor	-0.078	-0.183*	0.029	.
	Socioemotional	-0.038	-0.128	-0.086	.

THE UNIVERSITY OF CHICAGO

Table 5d: Correlation between Time to Full Mastery (At Midline/Endline Measurement Age) and Denver Endline Scores

		Time to Full Mastery			
		Language	Cognitive	Fine Motor	Gross Motor
Denver Score (Midline)	Language and Cognitive	-0.072	0.093*	0.001	0.150
	Fine Motor	0.045	-0.037	-0.051	0.062
	Gross Motor	0.012	0.015	-0.064	0.095
	Socioemotional	0.010	-0.029	0.013	0.006
Denver Score (Endline)	Language and Cognitive	0.118	0.027	-0.271**	.
	Fine Motor	-0.038	-0.008	-0.040	.
	Gross Motor	0.217	-0.027	-0.069	.
	Socioemotional	-0.174	-0.146	-0.167	.

- While all the correlations are in the expected direction, the different measures are far from perfectly correlated, suggesting that they capture different aspects of knowledge.
- Table 6 shows the correlations between different measures of knowledge.
- Time to first mastery is strongly negatively correlated with passing rates but much more weakly correlated with knowledge retention.
- Instability (backsliding) is at best weakly correlated with speed (time to mastery).
- The different measures of knowledge capture aspects of learning.

Table 6: Correlations between Different Measures of Knowledge

Correlation Variables	Language	Cognitive	Fine Motor	Gross Motor
Time to First Mastery vs. Avg. Passing Rate	-0.641***	-0.677***	-0.688***	-0.607***
Time to First Mastery vs. Instability	0.181***	0.208***	0.175***	-0.035
Avg. Passing Rate vs. Instability	-0.810***	-0.831***	-0.857***	-0.932***
Time to Full Mastery vs. Avg. Passing Rate	0.137***	0.193***	0.022	0.181***
Time to Full Mastery vs. Instability	0.170***	0.209***	0.253***	0.589***
Time to Full Mastery vs. Time to First Mastery	0.237***	0.155***	0.049*	-0.518***

THE UNIVERSITY OF CHICAGO

Stability of Mastery of Skills over Time

- Using our data and measures, we can define ability groups and determine the stability of membership in the ability categories. Ability categories are defined by the speed of mastering the task (time to the first correct answer).
- It is conventional to measure ability by the speed of learning, while learning is defined by eventual mastery of tasks. We examine how distinct these measures actually are.
- Table 7 defines the categories.

THE UNIVERSITY OF CHICAGO

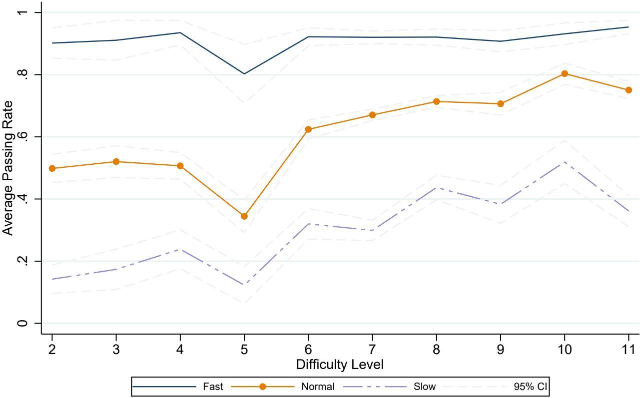
Table 7: Ability Categories (Measured over All Levels)

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

- There is strong persistence of passing rates across difficulty levels.
- Figure 5 shows that passing rates are persistent.
- Figures 6 and 7 show similar persistence for other measures of knowledge.
- The full mastery measure is quite noisy.
- Ability predicts the proportion of times that children get the wrong answer after a first correct answer (a measure of instability in performance) for cognition, language, and the other skills.

Figure 5: Average Passing Rate by Ability Category and Level

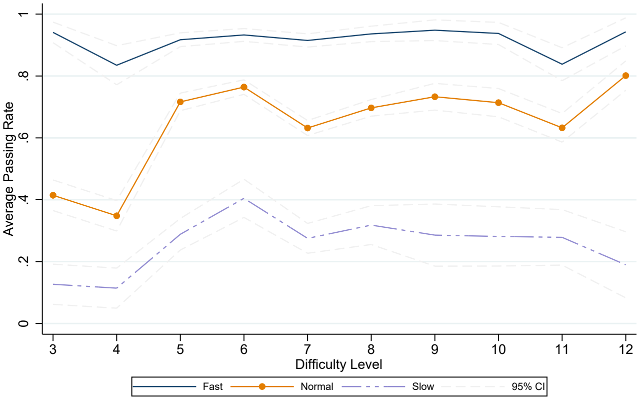
(a) Language Tasks



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

Figure 5: Average Passing Rate by Ability Category and Level, Cont'd

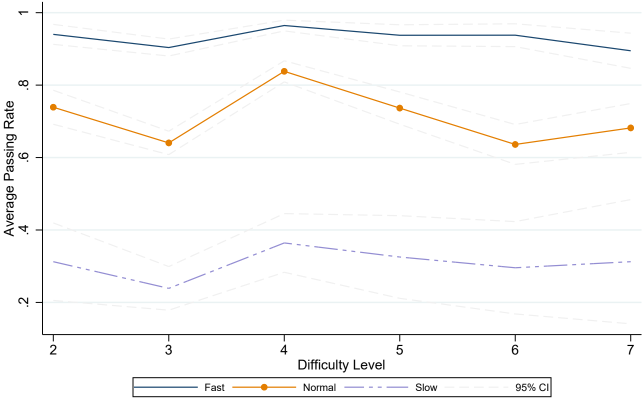
(b) Cognitive Tasks



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

Figure 5: Average Passing Rate by Ability Category and Level, Cont'd

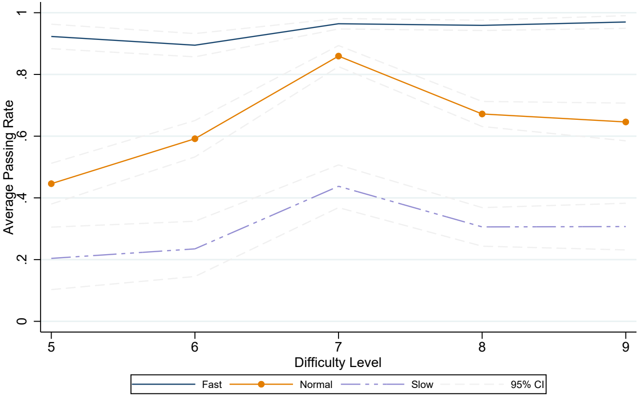
(c) Fine Motor Tasks



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

Figure 5: Average Passing Rate by Ability Category and Level, Cont'd

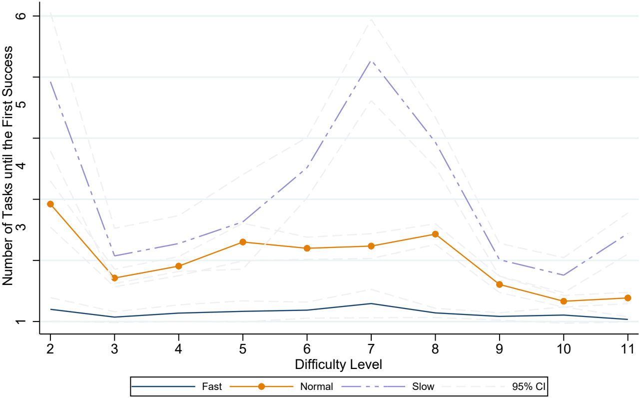
(d) Gross Motor Tasks



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

Figure 6: Time to First Mastery Measures by Ability Category and Level

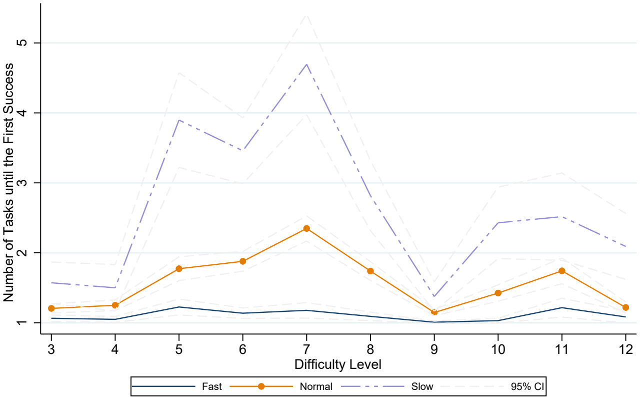
(a) Language Tasks



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

Figure 6: Time to First Mastery Measures by Ability Category and Level, Cont'd

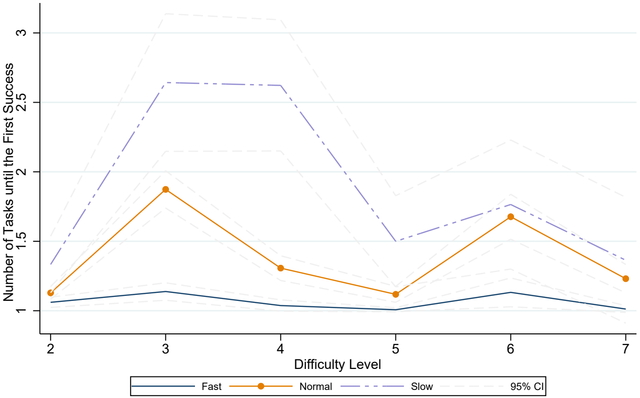
(b) Cognitive Tasks



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

Figure 6: Time to First Mastery Measures by Ability Category and Level, Cont'd

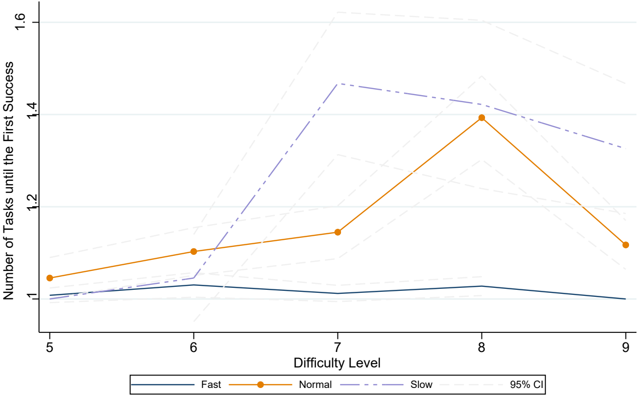
(c) Fine Motor Tasks



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

Figure 6: Time to First Mastery Measures by Ability Category and Level, Cont'd

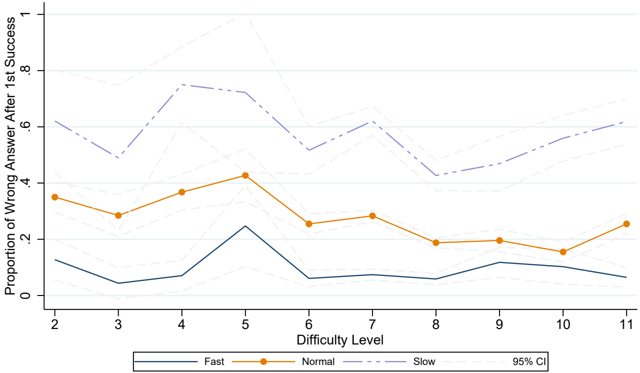
(d) Gross Motor Tasks



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

Figure 7: Instability (Proportion of Wrong Answers after First Success) Measures by Ability Category and Level

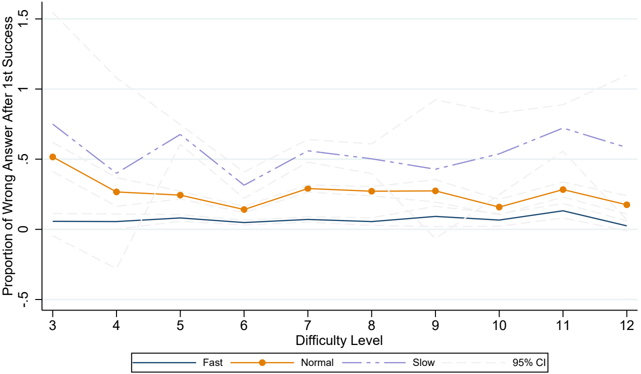
(a) Language Tasks



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

Figure 7: Instability (Proportion of Wrong Answers after First Success) Measures by Ability Category and Level, Cont'd

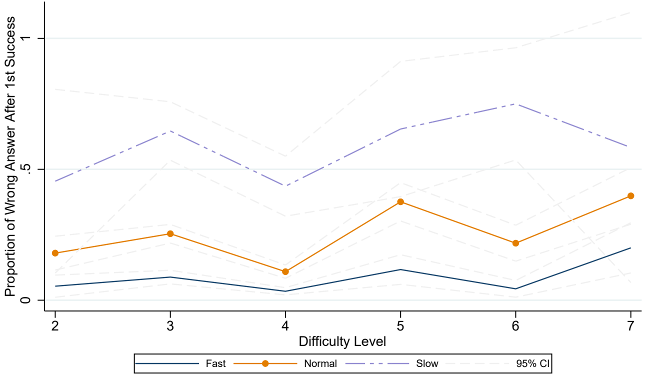
(b) Cognitive Tasks



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

Figure 7: Instability (Proportion of Wrong Answers after First Success) Measures by Ability Category and Level, Cont'd

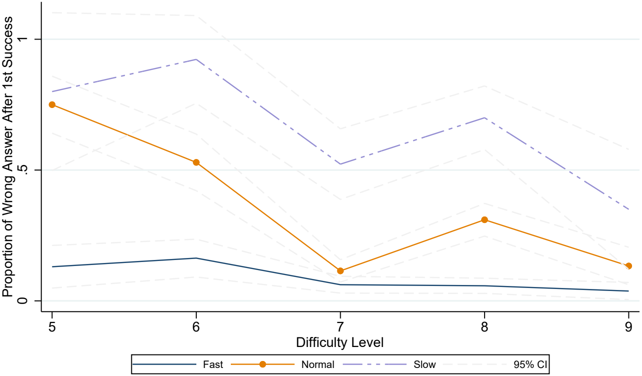
(c) Fine Motor Tasks



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

Figure 7: Instability (Proportion of Wrong Answers after First Success) Measures by Ability Category and Level, Cont'd

(d) Gross Motor Tasks



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

Testing Measured Skill Invariance

- Agostinelli and Wiswall (2021) raise important questions about the existence of invariant measures of skill.
- ***Mean measured skill invariance*** (our term) for measure $\boldsymbol{Z}(\boldsymbol{s}, \boldsymbol{a})$ of skill $\boldsymbol{s}$ at age $\boldsymbol{a}$ requires that

$$\boldsymbol{E}(\boldsymbol{Z}(\boldsymbol{s}, \boldsymbol{a}) \mid \boldsymbol{K}(\boldsymbol{s}, \ell, \boldsymbol{a}) = \tau) = \boldsymbol{E}(\boldsymbol{Z}(\boldsymbol{s}, \boldsymbol{a}') \mid \boldsymbol{K}(\boldsymbol{s}, \ell, \boldsymbol{a}') = \tau) \quad (3)$$

  for $\boldsymbol{a} \neq \boldsymbol{a}'$; that is, at the same ***true skill level*** τ, the measures of skill $\boldsymbol{s}$ at ages $\boldsymbol{a}$ and $\boldsymbol{a}'$ should coincide for all $\boldsymbol{a}, \boldsymbol{a}' \in [\underline{\boldsymbol{a}}(\ell), \bar{\boldsymbol{a}}(\ell)]$.
- To conduct this test, we need to find groups with the same latent skill levels $\boldsymbol{K}(\boldsymbol{s}, \ell, \boldsymbol{a})$ at different ages and then measure the child task performance $\boldsymbol{Z}(\boldsymbol{s}, \boldsymbol{a})$ for the different age groups.

Finding Groups with Same τ but Different a

- For all children in the intervention, we calculate average passing rates at each difficulty level for each skill throughout the entire intervention.
- To avoid small cells for our measures of knowledge, we array the data by quantiles of passing rates in the order of difficulty level.
- Table 8 uses passing rates on language skills at level ℓ and skill s-specific disaggregated UHP measures to test the condition $K(s, \ell, a) = K(s, \ell, a') = \tau$ (equal passing rates), a precondition for a test of measure invariance comparing age a and a' aggregated Denver scores.
- Based on the average passing rate at each difficulty level, we group the children with similar task performance together.

Table 8: Test of the Condition That $K(s, \ell, a) = K(s, \ell, a')$ for Language Skill Using UHP Difficulty Levels (Up to Denver Endline Age)

Level	Category	τ_1	τ_2	τ_3	τ_4
	Average Passing Rate				
	Young	0	0.283	0.723	1
	Old	0	0.321	0.656	1
	Test $K(s, \ell, a) = K(s, \ell, a')$: p-value		0.148	**0.004**	
	N	117	112	112	108
	Latent Skill Range	[0, 0]	[0.077, 0.5]	[0.5, 0.917]	[1, 1]
	Age at Enrollment (Months)				
2	Young	12.432	10.267	10.049	13.611
	Old	17.909	13.940	13.871	18.352
	Test $a = a'$: p-value	**0.000**	**0.000**	**0.000**	**0.000**
	Average Starting Age at Level 2				
	Monthly Age (Young)	13.186	10.543	10.179	14.676
	Monthly Age (Old)	19.103	13.991	14.478	20.000
	Curriculum Age Range for Level 2: [6.75, 20]				

Continues

THE UNIVERSITY OF CHICAGO

Table 8: Test of the Condition That $K(s, \ell, a) = K(s, \ell, a')$ for Language Skill Using UHP Difficulty Levels (Up to Denver Endline Age), Cont'd

Level	Category	τ_1	τ_2	τ_3	τ_4
	Average Passing Rate				
	Young	0	0.513	1.000	
	Old	0	0.514	1.000	
	Test $K(s, \ell, a) = K(s, \ell, a')$: p-value		0.969		
	N	122	136	134	
	Latent Skill Range	[0, 0]	[0.2, 0.8]	[1, 1]	
	Age at Enrollment (Months)				
3	Young	12.162	10.147	11.715	
	Old	17.140	13.866	16.480	
	Test $a = a'$: p-value	**0.000**	**0.000**	**0.000**	
	Average Starting Age at Level 3				
	Monthly Age (Young)	14.035	11.638	13.352	
	Monthly Age (Old)	17.671	15.310	17.286	
	Curriculum Age Range for Level 3: [9.5, 18.25]				

THE UNIVERSITY OF CHICAGO

Testing Measured Skill Invariance

- We next test the hypothesis that the aggregate Denver tests for s-comparable skills satisfy the criterion $E(Z(s,a) \mid K(s,\ell,a)=\tau) = E(Z(s,a') \mid K(s,\ell,a')=\tau)$ for different skills.
- Our Denver test endline measures are comparable to other commonly used achievement and assessment tests such as the Bailey tests.
- Tables 9a–9b report tests of whether the means of raw Denver scores are different (e.g., young vs. old) for each partition of τ at each difficulty level.

THE UNIVERSITY OF CHICAGO

Table 9a: Tests of the Mean Differences of Raw Denver Language Score $Z(s,a)$ Conditional on Language τ Groups by Difficulty Levels (Up to Denver Endline Age)

Denver	Category	τ_1	τ_2	τ_3	τ_4
UHP Language Level 2					
Endline	Young	26.271	24.306	24.447	26.486
	Old	29.956	28.056	28.159	29.237
(Language and Cognitive)	*p*-value	**0.000**	**0.000**	**0.000**	**0.004**
UHP Language Level 3					
Endline	Young	26.180	24.081	25.813	
	Old	28.786	28.191	27.957	
(Language and Cognitive)	*p*-value	**0.002**	**0.000**	**0.012**	
UHP Language Level 4					
Endline	Young	26.949	24.580	23.882	25.872
	Old	29.278	27.889	27.553	28.892
(Language and Cognitive)	*p*-value	**0.023**	**0.000**	**0.000**	**0.000**
UHP Language Level 5					
Endline	Young	24.966	23.940	25.250	
	Old	28.848	26.357	26.750	
(Language and Cognitive)	*p*-value	**0.000**	**0.000**	0.313	
UHP Language Level 6					
Endline	Young	29.323	25.467	25.440	27.385
	Old	32.321	30.427	30.292	31.742
(Language and Cognitive)	*p*-value	**0.011**	**0.000**	**0.000**	**0.000**

Table 9b: Tests of the Mean Differences of Raw Denver Language Score $Z(s, a)$ Conditional on Language τ Groups by Difficulty Levels (Up to Denver Endline Age)

Denver	Category	τ_1	τ_2	τ_3	τ_4	τ_5
	UHP Language Level 7					
Endline	Young	27.148	27.518	26.183	26.182	25.532
	Old	30.300	32.145	31.067	31.725	31.042
(Language and Cognitive)	p-value	**0.003**	**0.000**	**0.000**	**0.000**	**0.000**
	UHP Language Level 8					
Endline	Young	26.942	27.000	26.102	28.237	25.339
	Old	29.333	31.442	32.526	32.320	30.600
(Language and Cognitive)	p-value	**0.025**	**0.000**	**0.000**	**0.000**	**0.000**
	UHP Language Level 9					
Endline	Young	27.500	29.516	25.773		
	Old	31.525	32.247	30.615		
(Language and Cognitive)	p-value	**0.000**	**0.000**	**0.000**		
	UHP Language Level 10					
Endline	Young	25.579	28.048	30.756	27.692	
	Old	28.300	29.692	32.886	32.136	
(Language and Cognitive)	p-value	0.163	0.151	**0.005**	**0.000**	
	UHP Language Level 11					
Endline	Young	27.129	27.519	26.063		
	Old	30.609	32.218	31.072		
(Language and Cognitive)	p-value	**0.000**	**0.000**	**0.000**		

- We find that, for raw Denver scores, the old group's performance at the same level of measured knowledge is consistently better than the young group's performance; i.e., condition (3) is almost always violated, so the condition $\boldsymbol{E}(\boldsymbol{Z}(\boldsymbol{s}, \boldsymbol{a}) \mid \boldsymbol{K}(\boldsymbol{s}, \ell, \boldsymbol{a}) = \tau) = \boldsymbol{E}(\boldsymbol{Z}(\boldsymbol{s}, \boldsymbol{a}') \mid \boldsymbol{K}(\boldsymbol{s}, \ell, \boldsymbol{a}') = \tau)$ does not hold, even though the disaggregated measures of skill are the same.
- Measured skill invariance is rejected.
- Other factors beside pure knowledge of $\boldsymbol{s}$, as we measure it, affect Denver tests.

THE UNIVERSITY OF CHICAGO

Denver Language Test Results

- The previous tests report tests of hypothesis (3) using combined Denver language and cognitive tests.
- Scores are combined because there are few Denver test items for cognition.
- Our rejections for the Denver tests may be a consequence of these scores combining conceptually distinct skills.
- We conduct a similar series of tests using only language tests.
- In Tables 10a–10b, we continue to reject the skill invariance assumption for language skill even after only considering the Denver language items.

Table 10a: Tests of the Mean Differences of Raw Denver Language Score $Z(s, a)$ Conditional on Language τ Groups by Difficulty Levels (Up to Denver Endline Age)

Denver	Category	τ_1	τ_2	τ_3	τ_4
	UHP Language Level 2				
Endline	Young	22.229	20.652	21.463	22.405
	Old	24.622	23.976	22.789	24.026
(Language)	p-value	**0.000**	**0.000**	**0.009**	**0.011**
	UHP Language Level 3				
Endline	Young	22.220	20.774	21.958	
	Old	23.667	23.489	23.191	
(Language)	p-value	**0.012**	**0.000**	**0.032**	
	UHP Language Level 4				
Endline	Young	22.744	20.902	21.059	21.974
	Old	24.056	23.143	23.132	23.757
(Language)	p-value	**0.056**	**0.000**	**0.000**	**0.001**
	UHP Language Level 5				
Endline	Young	21.458	20.700	21.750	
	Old	23.909	22.167	22.500	
(Language)	p-value	**0.000**	**0.000**	0.455	
	UHP Language Level 6				
Endline	Young	24.387	21.987	21.713	22.949
	Old	26.536	25.123	24.623	26.097
(Language)	p-value	**0.009**	**0.000**	**0.000**	**0.000**

Table 10b: Tests of the Mean Differences of Raw Denver Language Score $Z(s, a)$ Conditional on Language τ Groups by Difficulty Levels (Up to Denver Endline Age)

Denver	Category	τ_1	τ_2	τ_3	τ_4	τ_5
	UHP Language Level 7					
Endline	Young	22.833	22.911	22.361	22.056	21.729
	Old	24.980	26.309	25.659	26.000	25.447
(Language)	p-value	**0.004**	**0.000**	**0.000**	**0.000**	**0.000**
	UHP Language Level 8					
Endline	Young	22.712	22.672	22.276	23.210	21.673
	Old	24.286	25.977	26.526	26.479	25.109
(Language)	p-value	**0.032**	**0.000**	**0.000**	**0.000**	**0.000**
	UHP Language Level 9					
Endline	Young	23.333	24.355	21.883		
	Old	25.750	26.476	25.198		
(Language)	p-value	**0.000**	**0.000**	**0.000**		
	UHP Language Level 10					
Endline	Young	21.842	23.698	24.953	23.154	
	Old	23.500	24.675	26.886	26.311	
(Language)	p-value	0.187	0.202	**0.003**	**0.000**	
	UHP Language Level 11					
Endline	Young	22.803	23.013	22.099		
	Old	25.217	26.385	25.505		
(Language)	p-value	**0.000**	**0.000**	**0.000**		

Robustness to Age of Entry

- A feature of China REACH is that all children of the same age are taught and examined on the same task.
- The late entrants have fewer lessons and may not be at the same level of knowledge due to dynamic complementarity of knowledge.
- However, we condition on knowledge $\boldsymbol{K}(\boldsymbol{s}, \ell, \boldsymbol{a})$ attained, so this consideration does not affect our analysis.
- Nonetheless, we conduct a series of robustness checks and find that our conclusions are not affected by alternative treatments of late entrants.

Conclusion

- This paper tests and rejects a key assumption invoked in the economics of education and in the analysis of skill formation: the existence of invariant measures of skill across different levels of the same skill (“human capital”).
- This assumption underlies a large body of research in the social sciences.
- Value-added measures are widely used to measure the output of schools.
- Aggregate test scores are used to measure gaps in skills across demographic groups.
- This paper shows that this practice is unwise.

- The aggregate measures used to chart student gains, child development, and the contribution of teachers and caregivers to student development are not comparable over time and persons except, possibly, for narrowly defined measures of skill.
- Accurate skill measurement requires much more disaggregated approaches, and conventional measures that assume invariance are fragile and should be used with caution if at all.

Testing Exchangeability Using Regressions

- We generate an indicator vector.
- For example, for the three-task case, we test the following:

$$\Pr(001) > \Pr(010) \quad (4)$$
$$\Pr(001) > \Pr(100) \quad (5)$$
$$\Pr(011) > \Pr(110) \quad (6)$$
$$\Pr(011) > \Pr(101) \quad (7)$$

- For each child $\boldsymbol{i}$, indicator vector of pattern $\boldsymbol{k}$, and difficulty level ℓ, we have the system of equations:

$$\boldsymbol{D}_{\boldsymbol{i}}^{\boldsymbol{k},\ell} = \boldsymbol{Z}'_{\boldsymbol{i},\boldsymbol{k},\boldsymbol{\ell}}\boldsymbol{\beta}^{\boldsymbol{k},\boldsymbol{\ell}} + \varepsilon_{\boldsymbol{i},\boldsymbol{k},\ell}. \tag{8}$$

- Table 11 illustrates the structure of our tests for patterns of three tasks (without controls).

Table 11: Hypothesis Tests for Patterns of Three Tasks (Without Controls)

	Learning Pattern		Random Pattern							
Level	Pattern	Pr(Pattern)	Pattern	Pr(Pattern)	Pattern	Pr(Pattern)	Null Hypothesis	Chi-square	*p*-value	df
2	001	0.571	010	0.429	100	0.000	Pr(001)=Pr(010)	0.144	0.704	1
	011	0.692	101	0.231	110	0.077	Pr(011)=Pr(101)=Pr(110)	10.233	**0.006**	2
3	001	0.714	010	0.048	100	0.238	Pr(001)=Pr(010)=Pr(100)	19.908	**0.000**	2
	011	0.640	101	0.200	110	0.160	Pr(011)=Pr(101)=Pr(110)	8.874	**0.012**	2
4	001	0.313	010	0.313	100	0.375	Pr(001)=Pr(010)=Pr(100)	0.118	0.943	2
	011	0.600	101	0.267	110	0.133	Pr(011)=Pr(101)=Pr(110)	5.330	**0.070**	2
5	001	0.545	010	0.273	100	0.182	Pr(001)=Pr(010)=Pr(100)	2.222	0.329	2
	011	0.333	101	0.000	110	0.667	Pr(011)=Pr(110)	1.053	0.305	1
6	001	0.391	010	0.348	100	0.261	Pr(001)=Pr(010)=Pr(100)	0.661	0.719	2
	011	0.527	101	0.327	110	0.145	Pr(011)=Pr(101)=Pr(110)	15.812	**0.000**	2
7	001	0.500	010	0.500	100	0.000	Pr(001)=Pr(010)	0.000	1.000	1
	011	0.667	101	0.333	110	0.000	Pr(011)=Pr(101)	0.357	0.550	1
8	001	0.833	010	0.000	100	0.167	Pr(001)=Pr(100)	3.243	**0.072**	1
	011	0.778	101	0.222	110	0.000	Pr(011)=Pr(101)	3.409	**0.065**	1
9	001	0.300	010	0.400	100	0.300	Pr(001)=Pr(010)=Pr(100)	0.183	0.913	2
	011	0.273	101	0.318	110	0.409	Pr(011)=Pr(101)=Pr(110)	0.615	0.735	2
10	001	0.250	010	0.500	100	0.250	Pr(001)=Pr(010)=Pr(100)	0.411	0.814	2
	011	0.636	101	0.273	110	0.091	Pr(011)=Pr(101)=Pr(110)	7.284	**0.026**	2
11	001	0.571	010	0.214	100	0.214	Pr(001)=Pr(010)=Pr(100)	5.619	**0.060**	2
	011	0.364	101	0.418	110	0.218	Pr(011)=Pr(101)=Pr(110)	4.311	0.116	2

THE UNIVERSITY OF CHICAGO

Table 12: Percentage of Tests within Each Level Rejecting the No Learning Hypothesis: Tests of Exchangeability

Level	Language	Cognitive	Fine Motor	Gross Motor
	Rejection Rates			
1	N/A	N/A	80%	N/A
2	100%	64.3%	N/A	N/A
3	100%	N/A	100%	N/A
4	100%	N/A	75%	100%
5	100%	54.5%	50%	N/A
6	100%	92.3%	100%	N/A
7	100%	90.9%	50%	50%
8	100%	92.9%		100%
9	88.9%	N/A		
10	100%	66.7%		
11	100%	77.8%		
12		50%		
Overall	98.6%	77.9%	84.2%	83.3%

THE UNIVERSITY OF CHICAGO

Learning Pattern Features (Heterogeneity and State Dependence)

Model Descriptions

Model 1 (Probit Model): Polya Urn 1

- This model assumes no learning.
- For each task at a given difficulty level, the latent process is as follows:

$$Y_i^*(t) = \boldsymbol{X}'\beta + \varepsilon_{it}, \quad E(\varepsilon_{it}) = 0 \tag{9}$$
$$\varepsilon_{it} \perp\!\!\!\perp \boldsymbol{X} \quad \forall t \; (\varepsilon_{it} \text{ independent of } \boldsymbol{X}\, \forall t),$$

where $Y_i(t)^*$ is the latent value of the child i of the task t.

- ε_{it} is i.i.d. across individuals and tasks, so there is no persistent heterogeneity of ability.

THE UNIVERSITY OF CHICAGO

- $\boldsymbol{Y}_i(\boldsymbol{t})$ takes the value of zero when the child cannot pass the task $\boldsymbol{t}$:

$$\boldsymbol{Y}_i(\boldsymbol{t}) = \begin{cases} 1 & \boldsymbol{Y}_i^*(\boldsymbol{t}) \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

- This is a Bernoulli model with heterogeneity arising from observables.

Model 2 (Heterogeneity): Polya Urn 2

- Model 2 introduces an unobserved (by the analyst) individual effect that persists over trials but does not allow for learning.
- That is, for each task $\boldsymbol{t}$ at a given difficulty level, we have the following:

$$Y_i^*(\boldsymbol{t}) = \boldsymbol{X}'\boldsymbol{\beta} + \theta_i + \varepsilon_{it}. \tag{10}$$

$$Y_i(\boldsymbol{t}) = \begin{cases} 1 & Y_i^*(\boldsymbol{t}) \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where θ_i is the individual-specific latent factor, which has mean zero and variance σ_θ^2 and is independent of $\boldsymbol{X}$.

Model 3 (State Dependence: Learning): Polya Urn 3

- Model 3 is a model of true state dependence, which captures learning.
- It can be represented as follows:

$$Y_i^*(t) = X'\beta + \delta \sum_{k=1}^{t-1} Y_i(k) + \varepsilon_{it}. \tag{11}$$

$$Y_i(t) = \begin{cases} 1 & Y_i^*(t) \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where ε_{it} is i.i.d. with mean zero, the latent value $Y_i(t)^*$ depends on the past task performance $\{Y_i(k)\}_{k=1}^{t-1}$, and ε_{it} is independent of X.

- We use $\sum_{k=1}^{t-1} \boldsymbol{Y}_i(\boldsymbol{k})$ as a measure of performance on previous tasks.
- This is one way to capture the notion that success produces success.

Model 4 (Heterogeneity and State Dependence): Combine Polya Urns 2 & 3

- Model 4 is a state dependence model with individual unobserved heterogeneity.
- This model can be written as:

$$\boldsymbol{Y}_i^*(\boldsymbol{t}) = \boldsymbol{X}'\boldsymbol{\beta} + \delta \sum_{\boldsymbol{k}=1}^{\boldsymbol{t}-1} \mathbf{Y}_i(\boldsymbol{k}) + \theta_i + \varepsilon_{\boldsymbol{it}}. \tag{12}$$

$$\boldsymbol{Y}_i(\boldsymbol{t}) = \begin{cases} 1 & \boldsymbol{Y}_i^*(\boldsymbol{t}) \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

- As described previously, ε_{it} is i.i.d. with mean zero, and independent of $\boldsymbol{X}$, θ_i, and $\sum_{k=1}^{t-1} \boldsymbol{Y}_i(\boldsymbol{k})$.
- The latent value $\boldsymbol{Y}_i(\boldsymbol{t})^*$ depends on the cumulative past task performance $\{\boldsymbol{Y}_i(\boldsymbol{k})\}_{k=1}^{t-1}$ and individual heterogeneity.

Model 5 (State Dependence with a Proxy for Ability Duration): Polya Urn 4

- Model 5 is a model of state dependence that adds the time to mastery measure at previous difficulty levels as a proxy for ability.
- This model can be written as:

$$\boldsymbol{Y}^*_{i,\ell}(\boldsymbol{t}) = \boldsymbol{X}'\boldsymbol{\beta} + \delta \sum_{k=1}^{t-1} \boldsymbol{Y}_{i,l}(\boldsymbol{k}) + \gamma \boldsymbol{D}_{i,\ell-1} + \varepsilon_{it}, \tag{13}$$

where ε_{it} is i.i.d. with mean zero, and the latent value $\boldsymbol{Y}_{i,\ell}(\boldsymbol{t})^*$ at difficulty level ℓ depends on past task performance at the same level $\{\boldsymbol{Y}_{i,\ell}(\boldsymbol{k})\}_{k=1}^{t-1}$.

- $D_{i,\ell-1}$ represents the number of attempts required to get the first correct answer at the previous difficulty level $\ell - 1$.
- It is a measure of ability and captures the children's heterogeneity.

$$Y_{i,\ell}(t) = \begin{cases} 1 & Y^*_{i,\ell}(t) \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

THE UNIVERSITY OF CHICAGO

Model 6 (Current and Lagged State Dependence): A Version of Polya Urn 4

- Model 6 is a model of state dependence with individual unobserved heterogeneity.
- The difference between model 4 and 6 is that model 6 also includes the product term that reflects state dependence (i.e., $\delta_2 \sum_{j=1}^{t-1} \prod_{m=1}^{j} \boldsymbol{Y}_{i,\ell}(\boldsymbol{t} - \boldsymbol{m})$).
- This is an indicator of the number of correct answers up to that point.
- It is a renewal process (length of current streak of successful answers).

THE UNIVERSITY OF CHICAGO

- This model may be written as:

$$\boldsymbol{Y}^*_{i,\ell}(\boldsymbol{t}) = \boldsymbol{X}'\boldsymbol{\beta} + \delta \sum_{k=1}^{t-1} \boldsymbol{Y}_{i,\ell}(\boldsymbol{k}) + \delta_2 \sum_{j=1}^{t-1} \Pi_{m=1}^{j} \boldsymbol{Y}_{i,\ell}(\boldsymbol{t}-\boldsymbol{m}) + \theta_i + \varepsilon_{it}, \tag{14}$$

where ε_{it} is i.i.d. with mean zero, and the latent value $\boldsymbol{Y}^*_{i,\ell}(\boldsymbol{t})$ at difficulty level ℓ depends on past task performance at the same level $\{\boldsymbol{Y}_{i,\ell}(\boldsymbol{k})\}_{k=1}^{t-1}$ as well as individual heterogeneity.

- It is independent of θ, $\boldsymbol{X}$, and all $\boldsymbol{Y}_{i,\ell}(\boldsymbol{t}-\boldsymbol{m})$ for $\boldsymbol{m} > \delta$.

$$\boldsymbol{Y}_{i,\ell}(\boldsymbol{t}) = \begin{cases} 1 & \boldsymbol{Y}^*_{i,\ell}(\boldsymbol{t}) \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Comparing Model Fits

Table 13: Fine Motor Skill Level 1 Cases with Three Tasks χ^2 Test

		Predicted Number				
Pattern	Observation	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect + State Dependence I	Model 6 Random Effect + State Dependence II
000	3	1.629	2.426	2.080	2.026	2.161
001	7	3.549	3.299	5.070	4.017	3.539
010	5	3.549	3.299	3.692	3.223	4.350
100	2	3.549	3.299	2.974	2.829	3.116
011	8	9.551	7.505	12.503	10.581	11.282
101	4	9.551	7.505	9.337	8.231	6.263
110	6	9.551	7.505	6.740	6.182	9.222
111	37	31.072	37.163	29.603	34.912	32.067
χ^2		11.711	7.649	8.526	6.841	7.864
Theoretical χ^2 at 5%		14.067	14.067	14.067	14.067	14.067
p-value		0.110	0.365	0.289	0.446	0.345
D.F.		7	7	7	7	7

THE UNIVERSITY OF CHICAGO

Table 14: Percentage of Chi-Squared Test Not Rejecting the Null Hypothesis of Task Performance Patterns

Skill	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect + State Dependence I	Model 5 Duration + State Dependence	Model 6 Random Effect + State Dependence II
Language	37.3%	41.1%	66.1%	75.4%	24.4%	70.9%
Cognitive	38.8%	56.3%	52%	71.4%	30.2%	73.7%
Fine Motor	47.8%	54.5%	56.5%	73.9%	41.2%	75%
Gross Motor	36.3%	50%	50%	88.9%	0%	66.7%

Table 15: Percentage of Smallest BIC across Models

Skill	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect + State Dependence I	Model 5 Duration + State Dependence	Model 6 Random Effect + State Dependence II
Language	8.2%	4.9%	16.4%	0.0%	70.5%	0.0%
Cognitive	2.0%	2.0%	70.0%	0.0%	26.0%	0.0%
Fine Motor	0.0%	0.0%	73.9%	0.0%	26.1%	0.0%

Return to “Backsliding” slide

THE UNIVERSITY OF CHICAGO