Tests of Early Childhood Skill Learning Patterns*

Jin Zhou, James Heckman, Fuyao Wang, and Bei Liu

Center for the Economics of Human Development
University of Chicago
September 1, 2021

Abstract

This paper uses novel experimental data from a prototypical home visiting program in China with high-frequency measurements to investigate the growth of multiple skills. After identifying the existence of multiple skill development during the intervention, we further study the features of the children learning patterns. We find that individual heterogeneity and previous task performance (state dependence) are key properties of the child's task performance in the intervention.

JEL Codes: J1 D83 C1 C5

Keywords: Child development, Testing, Learning

*This research was supported in part by: NIH grant NICHD R37HD065072, and the Institute for New Economic Thinking and an anonymous donor. We thank our partners on the China REACH project: China Development Research Foundation, Child Development Research Group at the University of West Indies, and Dr. Sally Grantham-McGregor. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders, partners, or the official views of the National Institutes of Health. Erlfang Tsai, and Alejandra Campos contributed highly competent and insightful research assistance and commentary.

Jin Zhou
Center for the Economics
of Human Development
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Email: jinzhou@uchicago.edu

James J. Heckman
Center for the Economics
of Human Development
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Email: jjh@uchicago.edu

Fuyao Wang
Center for the Economics
of Human Development
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Email: fywang@uchicago.edu

Bei Liu
China Development Research
Foundation
Floor 15, Tower A
Imperial International Center
No. 136, Andingmen Wai Avenue
Dongcheng District
Beijing, P.C. 100011
Email: liubei@cdrf.org.cn

1 Introduction

A growing body of research establishes the effectiveness of home visiting programs targeted to the early years in developing the skills of disadvantaged children (see, e.g., Howard and Brooks-Gunn, 2009; HomVEE, 2020; Grantham-McGregor and Smith, 2016). They are relatively of low cost compared to many other early childhood programs. They place minimal demands on the training required of the home visitors and on the infrastructure needed to support them. Visitors have levels of education comparable to those of the caregivers visited. The Jamaica Reach Up and Learn program, established some 30 years ago, is a successful prototype of a home visiting program emulated worldwide.

This paper studies a close replica of the original Jamaica Reach Up and Learn program, China REACH, which was brought to scale in an impoverished region of Western China (1500+ participants compared to the 100+ participants in the original Jamaica study). Zhou et al. (2021) show that the program can be successfully implemented at scale. There are two goals of this paper: the first goal is to study whether the children learn and develop their multiple skills, and the second goal is to examine the best properties of child learning pattern during the intervention.

The China REACH program gathers week-by-week growth data that make it possible for us to analyze the two questions stated above: first, whether the growth of multiple skills exists, and second the characteristics of the children's skill formation during the intervention. In the study, we find clear empirical evidence on the existence of learning patterns based on the children's task performance. That is when the children are taught similar tasks on multiple occasions, the probability of mastering the task increased. We use linear probability models to compare the learning patterns of children (i.e., Heckman, 1978). We find substantial evidence of learning.

We further examine the properties of the learning process during the intervention. We use six models with different properties: individual heterogeneity, state dependence, and approximate ability measures to test model fit. Heterogeneity recognizes that people are intrinsically different and learn at different rates. State dependence captures the growth of skills after instruction. Both processes are at work.

The paper unfolds in the following way. Section 2 sketches the China Reach Program design and unique measures. Section 3 tests for the existence of skill formation during the intervention. Section 4 formalizes the evidence of two features of child task learning. Section 5 concludes.

2 The Background of China REACH Program and Its Measures

The China REACH (Huachi) project was launched in 2015 by the China Development Research Foundation to improve social mobility in rural areas. It is a large scale program evaluated by a randomized control trial (RCT) designed to evaluate the impact of a home visiting delivery model for disadvantaged families in rural China.[1] It is based on a successful Jamaican prototype for disadvantaged children (e.g.,

[1] Huachi County in Gansu Province, one of the poorest areas in China, is the first site for the China REACH intervention.

Grantham-McGregor and Smith, 2016; Gertler et al., 2014, 2021).[2] The China REACH program aims to improve the health and multiple skill development of children by enhancing their engagement with caregivers and the larger community.

2.1 The Program

The program uses home visitors with educational attainments similar to those of the caregivers visited (e.g., most of them with nine years of mandatory education). College-educated psychologists and educators trained supervisors with at least associates degrees. The supervisors train home visitors and organize weekly meetings with them to prepare for the next week's home visits and correct problems that arise when conducting previous home visits. They also assess the quality of home visitors and their visits using a predesigned and standardized tool developed by the program.

The program encourages caregivers to interact with their children in developmentally appropriate ways. Visitors engage the households weekly and provide one-hour parenting or care-giving guidance lessons and support.[3] In each home visit, the home visitor records information about caregiver engagement (e.g., who worked with the child during the visit, whether the home visitor taught the caregivers relevant tasks if the child could not participate in the home visit, who played with the child after the visit and with what frequency), and child performance (e.g., the tasks taught in

[2]The Jamaica program substantially raised employment, IQ, wages, and education. See Gertler et al. (2021).

[3]The protocols are based on those used by the Jamaica program, adapted to Chinese culture (e.g., changing songs into popular Chinese songs, adding the background of pictures, which are familiar to Chinese people). The protocol for children younger than 18 months old focuses on motor and language skill training. After 18 months old, the protocol adds more cognitive skill content (e.g., classification, pairing, and picture puzzles).

the last week, and new tasks in the current week). The home visitors encourage the caregiver to interact with the child warmly and to create strong attachments.

2.2 The Curriculum

The curriculum was designed to develop a child's cognitive, language, and psycho-social skills. The activities included learning the environment through labeling, describing objects and actions, responding to the child's vocalizations and actions, playing educational games, and using picture books and songs that facilitated language acquisition. The first 18 months of the intervention are based on Piagetian concepts such as the use of tools and object permanence (Uzgiris and Hunt, 1975). After 18 months, concepts such as size, shape, quantity, color, and classification based on Palmer (1971) were included. Particular emphasis is placed on the use of praise and giving positive feedback to both the mother and child. Each session's curriculum is designed based on the general child development process to ensure the activities are at the appropriate level for the child. Home visitors introduce tasks based on the curriculum's prescribed activities for each week. In the intervention, children at the same weekly age get the same lesson.

Understanding of each set of weekly tasks is assessed at the end of each session. Tasks are assigned in a stepwise fashion. A group of tasks of equal difficulty is assigned at the first step. Then the child is taken to the next difficulty level, where a series of equally difficult tasks are executed and assessed. The weekly assessment is a binary calculation of whether or not the child can complete the tasks. Level-by-level, we can determine the child's knowledge of performing that task. Each week the child is exposed to multiple skills. Unlike general assessments often used to monitor learning, our assessments are specific to the tasks being learned. The tasks and the assessments are closely synchronized. This produces tangible, scale-free measures of knowledge within difficulty levels. We chain performances across difficulty levels but do not claim that we can cardinalize knowledge across those levels.

On average, in each week, 3 to 4 different skills are taught. Each lesson targets a specific set of skills (i.e., one of the gross motor, fine motor, language, and cognitive).[4] The intervention strictly followed the design in the curriculum.[5] Neither the home visitor nor the caregiver control the lesson content. Lessons given are thus exogenous or "fixed" for each child. The curriculum followed over the full 36 months is displayed in the timeline figures in Heckman and Zhou (2021).

Figure 1 summarizes the categories of skills taught by age. Before 18 months, the intervention concentrates more on language and motor skills. After 18 months, the curriculum focuses more on cognitive skill development. Each of the four primary skills is further broken down into smaller subsections. For example, fine motor skills include two main subjects: one is the skill related to finger movements, i.e., grasp, release, and stitch; the other is related to drawing ability. Cognitive skill lessons include spatial skill, understanding objects, and objects' function; matching; ordering; and numbering skills.

[4] In the curriculum, there are 67 fine motor skill lessons, 22 gross motor skill lessons, 230 cognitive skill lessons, and 147 language skill lessons.

[5] In our data, more than 97% of home visit records are exactly matched with the child's weekly age. Neither the parents nor the home visitor can change the teaching content of the intervention.

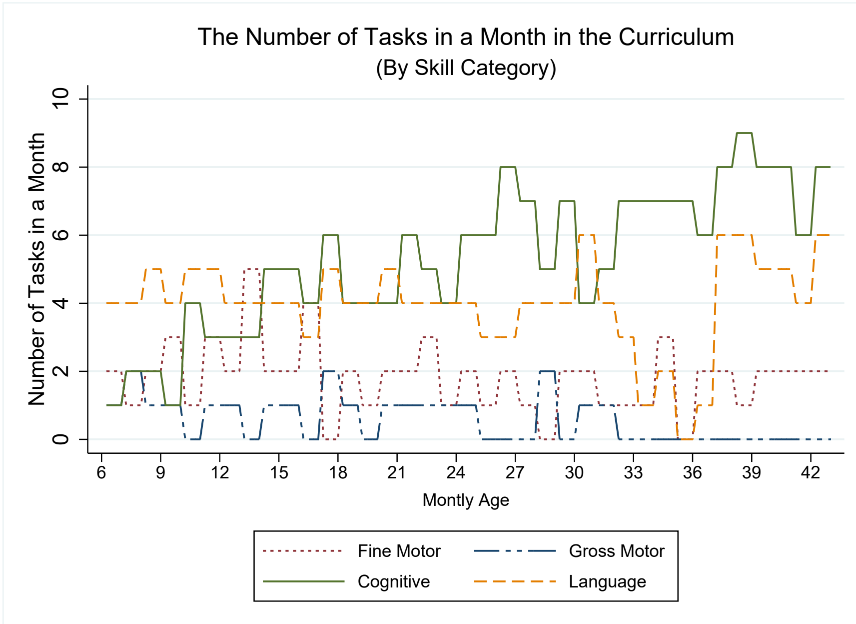


Figure 1: Curriculum Task Intensity by Skill Category

2.3 Skills Taught in the Curriculum

Gross motor skills, fine motor skills, language skills, and cognitive skills are taught. Within each skill group, based on the content in the skills, skills are ordered by difficulty level following the patterns developed by Palmer (1971). Skills are sorted into different difficulty levels. For example, for fine motor drawing lessons, there are 7 difficulty levels.[6,7] In general, the higher difficulty level for skills includes new content. For example, difficulty level 2 is to mimic circles. The skills at difficulty level 3 include drawing straight lines. We document how the tasks into different

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

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

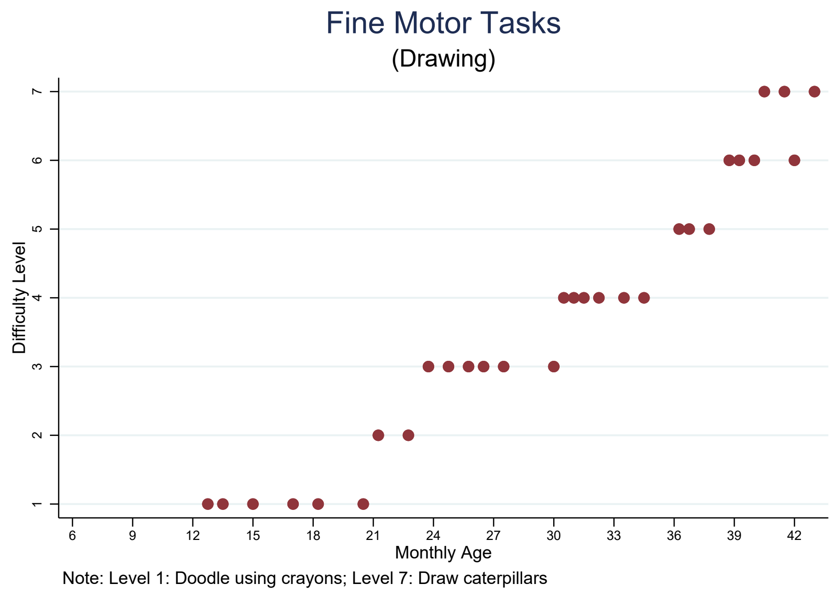
difficulty levels are categorized. See Appendix B in Heckman and Zhou (2021).

For example, Fine Motor Drawing lessons focus on a child's ability to use a writing utensil with increasing skill. First, a child is asked to hold the utensil to make markings. Next, the child should incorporate more and more cognitive skills to complete the tasks. They then begin by copying markings made by an adult. As skill levels progress, they are asked to make the marking after only a verbal command from the adult. Finally, the child progresses from abstract shapes to representative drawings. (See Table B3 in Heckman and Zhou (2021).)

In addition to tasks of different difficulty levels, the curriculum features multiple lessons and assessments at the same difficulty level l. The difficulty level category descriptions are listed in Appendix Curriculum. The number of lessons within each difficulty level depends on the content in the curriculum. For example, there are six assessments at difficulty level 3 for fine motor drawing skills and only 2 assessments at difficulty level 2.

Figure 2 gives the timing of each fine motor drawing assessment appearing in the curriculum design. In the figure, we see that for the designated skill, difficulty level 1 starts at 12 months and 3 weeks and lasts until 20 months and 2 weeks content. This timing means that when the child is 12 months and 3 weeks old, the home visitor will teach the first fine motor drawing skill to her. When she is 20 months and 2 weeks old, the home visitor will teach her the 6th lesson at difficulty level 1. In general, higher difficulty levels appear at later weekly ages. However, there can be some overlap across difficulty levels. For example, in Figure 2, by the time difficulty level 7 of fine motor lessons start, the last lesson of level 6 remains unfinished. In Figure 2, when fine motor lessons at difficulty 7 start, the student still receives lessons at difficulty level 6. Circling back is a strategy designed to solidify a child's understanding of a concept.

Figure 2: The Distribution of Fine Motor Skill Tasks Across Difficulty Levels



Another example concerns cognitive skill development. The curriculum divides cognitive skills into three categories: spatial awareness, knowledge of objects, and object functions. The difficulty levels in this domain are defined based on the abstract concepts, shown in Table 1. 74 discreet lessons are sorted into the listed 13 difficulty levels.[8] It covers the process of how the child learns to know an object and understand the function of the object.

The lessons in the cognitive knowledge of objects unit progress from a simple

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

understanding of the concept of pictures by acknowledging with vocalizations, to using receptive (heard) language to identify certain pictures. Receptive language, or the ability to understand language, is a skill developed prior to expressive language where a child forms words to communicate. The child must use his or her expressive language to complete the following lessons, which increase with difficulty as he or she must develop more and more language to identify an increasing number of images. To progress through level 7 and beyond, the child must display an increasingly sophisticated understanding of the stories presented, first simply naming actions, then answering questions, then talking abstractly about the story. Levels 10, 11, 12, and 13 ask the child to take the information presented and build on it by discussing the uses of objects presented and making connections with other images.

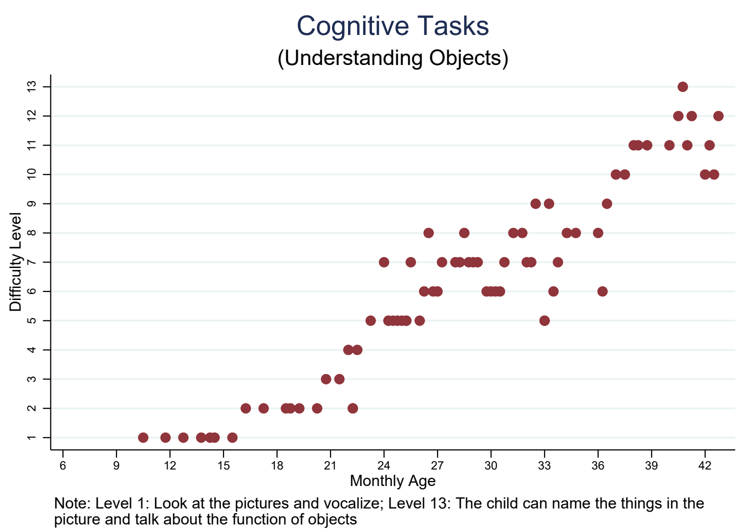


Figure 3: The Distribution of Cognitive Skill Tasks across Difficulty Levels

Figure 3 shows the timing of each cognitive level (knowing objects and understanding

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

Level 1	Look at the pictures and vocalize
Level 2	Name the objects and ask the baby to point to the pictures accordingly
Level 3	The child can name the objects in one picture, points to the named picture
Level 4	The child can name the objects in two or more pictures, points to the named picture
Level 5	The child can point out named pictures, says names of three or more
Level 6	The child can point out the picture mentioned and correctly names the name of 6 or more pictures
Level 7	The child can talk about the pictures, answer questions, understands, 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

the object's function) in the curriculum. The number of lessons varies across difficulty levels according to the curriculum content itself. Table 2 presents detailed information about the six lessons (and assessments) that are labeled as difficulty level one directed to ten-month to 15 month-old curriculum content. In Table 2, all lessons relate to the activity of looking at the pictures or objects and vocalizing, which does not require the child to name or identify the object.

Table 2: Cognitive Skill Task Content: Look at the pictures and vocalize (Level 1)

Difficulty Level	Month	Week	Learning Materials	Content
1	10	2	Picture book A	baby makes sound when looking at the pictures
1	11	3	Picture book B	baby looks at the pictures and vocalize
1	12	3	Picture book A	child makes sound looking at the pictures
1	13	3	Picture book B	child makes sound looking at the pictures
1	14	1	Picture book A	mother and child look at the pictures together, and let the child vocalize and touch the pictures
1	15	2	Picture book B	mother and child look at the pictures together, and let the child vocalize and touch the pictures

To summarize, the curriculum targets lessons at different skill levels for multiple levels of skill for each weekly age. For each type of skill, the task difficulty levels are based on the content of the tasks and the guideline of Uzgiris and Hunt (1975) and Palmer (1971) are constructed. The terms of the number of lessons within each difficulty level vary. We follow these scholars and assume that each level is a quantum of understanding comparable across children. We use achievement at each level of skill as our measure of knowledge.

3 Evidence on Learning

As documented above, each child was taught a few tasks within one difficulty level in the curriculum. The questions are whether we can identify the presence of children's skill growth based on the recorded task performance and, if so, how to explain children's growth of multiple skills.

3.1 Increasing passing rates

This section documents the children's task performance (i.e., average passing rate) for the tasks with each difficulty level. To examine whether skill growth exists, we first study the average passing rate at each difficulty level. Then in next section, following Heckman (1978), we examine the test exchangeability to compare the probability of "learning pattern" to the probability of "random patterns" in data. Following Heckman (1978, 1981), we examine whether the pattern of passing is consistent with chance guessing or systematic learning. The ideas is that, within each learning or internal passing is an exchangeable series of random variables (so the probability of passing is the same whatever the order of the tests) or if there is a pattern to learning that is not simply a function of age (maturation and learning from parents and peers). If the program adds to learning, the age-adjusted passing indicators should be exchangeable.[9]

First, we find a clear pattern that children's passing rates of tasks within one difficulty level are monotonically increasing. To provide a more intuitive idea, we first present some examples below.

As a first example, let's look at difficulty level 4 of the language skill tasks. In Figure 4, we plot the passing rates of tasks at level 4 by the order in which the children learn them. (e.g., 3 on the x-axis refers to the 3rd task the child learns at difficulty level 4.) In this figure, we can see that the passing rates are monotonically increasing.

[9]This test is motivated by empirical tests suggested in Bush and Mosteller (1955) and Restle and Greeno (1970), but the formal test is new due to Heckman (1978, 1981).

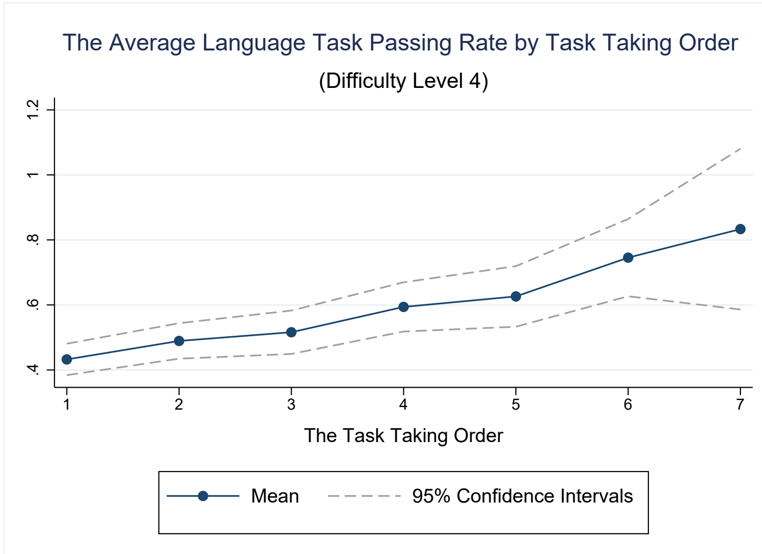


Figure 4

As a more robustness check, Table 3 tests whether the passing rates for later tasks are higher than the earlier tasks. This table shows that the passing rates of task three through seven are significantly higher than that of the first task. Similarly, when comparing to the second task, we can find that the passing rates are significantly higher from the fourth to the seventh tasks. The significant difference demonstrates that the more a child learns similar tasks at the same difficulty level, the better the child performs.

Table 3: Test of Language Task Passing Rate by Task Order (Difficulty Level 4)

Order	1	2	3	4	5	6	7
Mean	0.432	0.489	0.516	0.594	0.626	0.745	0.833
Standard Error	(0.025)	(0.028)	(0.034)	(0.038)	(0.047)	(0.059)	(0.112)
p value							
Test vs. Task 1		0.126	**0.046**	**0.000**	**0.000**	**0.000**	**0.006**
Test vs. Task 2			0.541	**0.028**	**0.014**	**0.000**	**0.019**
Test vs. Task 3				0.129	**0.061**	**0.002**	**0.032**
Test vs. Task 4					0.597	**0.044**	0.102
Test vs. Task 5						0.129	0.157
Test vs. Task 6							0.525

Bold values represent the $p < 0.1$

To present an overall picture of the increasing trend of the language skill, we show the average task passing rates by task order for all difficulty levels in Figure 5. The yellow lines in the figure indicate the last tasks at each difficulty level. We find that there exists the pattern that the passing rates are increasing within one level (the space between two yellow lines). Similarly, Figures 6-8 show such overall pictures for cognitive and fine motor. The same cannot be said of gross motor skills. In the first three figures, we see a similar pattern of increasing passing rates within each difficulty level. Thus, we infer from the aggregate task passing rate that the children's skill growth exists in our data. Next, to further investigate the learning behaviors, we study children's task performance at each level by comparing the probability of children showing the "learning pattern" with that of "random patterns".

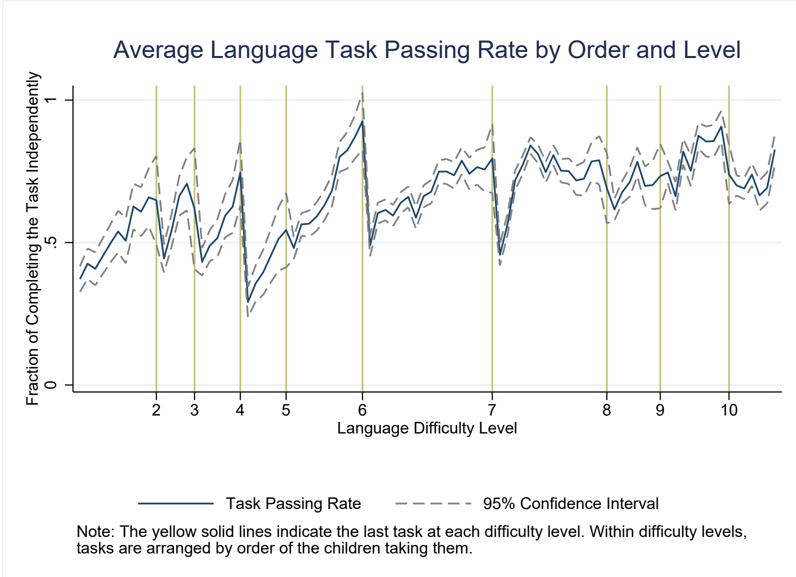


Figure 5

Figure 6

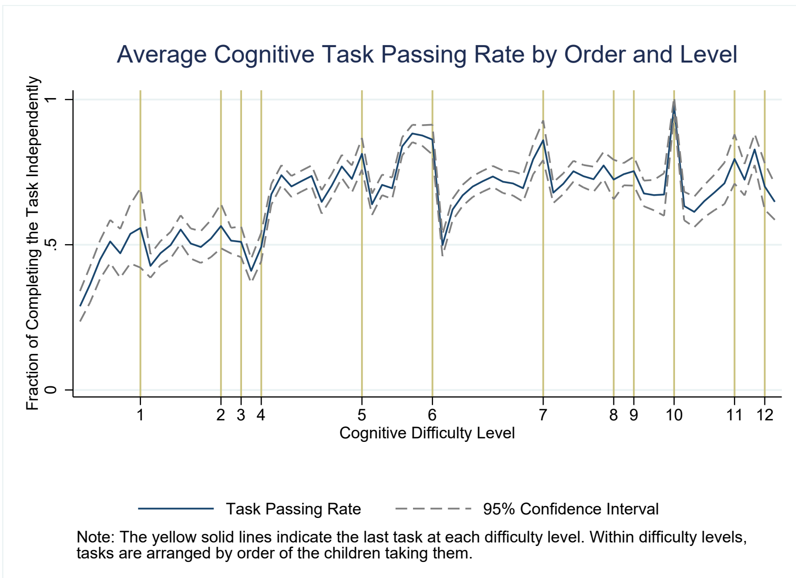


Figure 7

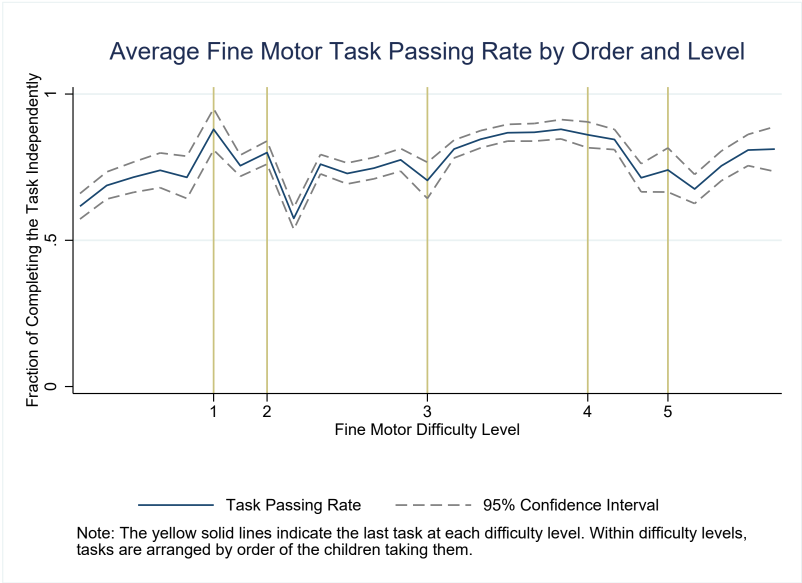
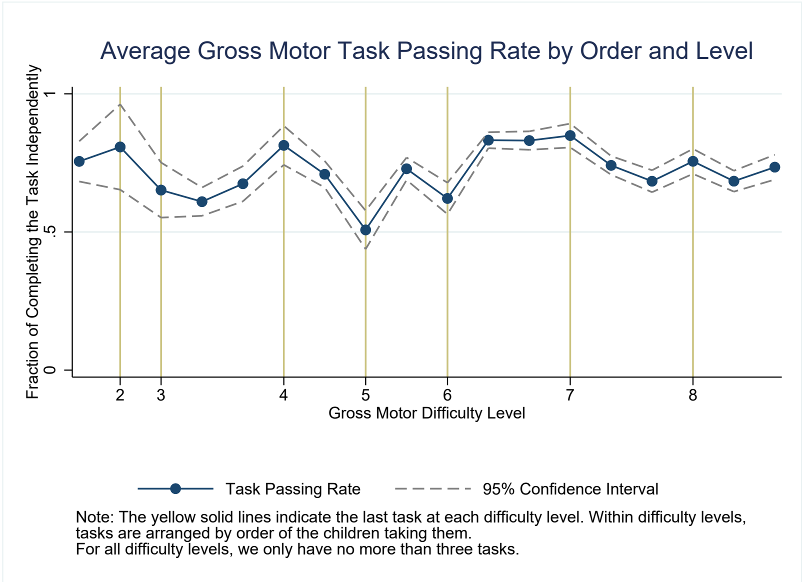


Figure 8



3.2 Test for Exchangeability Using Regressions

In this section, to diagnose whether the children in the program develop their skills, we want to examine whether the probability of learning pattern is higher than that of non-learning patterns (named as "random" patterns) for all four types of skills across all difficulty levels.

The learning pattern is defined as: at a given difficulty level, the child may fail to complete the initial task(s), but once the child completes a certain task independently, she (he) will never fail a task again at the same difficulty level. For example, in a series of 3 tasks, 001 and 011 are learning patterns, while 010, 100, 101, 110 are not.[10]

We follow Heckman (1978) and use linear probability models to compare the probability of a child showing the learning pattern to that of other non-learning patterns. The logic is that we generate an indicator vector for each week-by-week task performance pattern, run regressions simultaneously with the indicators as the dependent variables, and then test the equality of the coefficients of the regressions across patterns. For example, for the three task case, we will test the following:

[10]Here, we do not consider all passes and all fails cases.

$$
\begin{aligned}
\Pr(001) &> \Pr(010) \quad (1) \\
\Pr(001) &> \Pr(100) \quad (2) \\
\Pr(011) &> \Pr(110) \quad (3) \\
\Pr(011) &> \Pr(101) \quad (4)
\end{aligned}
$$

Here, "1" indicates the child completes the task independently (passes), and 0 means the child cannot complete the task independently (fails). The order refers to the task order the child took in the curriculum. The comparison is that among the cases with the same number of passed tasks, we compare the probability of "learning pattern" to the probability of "random pattern".

Using the three task case as an example, given the situation that the child only passed one task, we have three different patterns "001"(learning pattern), and "100", and "010" (random patterns). The test is to examine whether the equations (1) and (2) hold. Similarly, in the case that the child passed two tasks, we will test whether the equations (3) and (4) hold.

In the example, these are six different patterns for the case with three tasks, there are six indicator vectors D_i^k which correspond to each pattern. For each child i, and the indicator vector of the pattern k for difficulty level l, we have the system of equations:

$$D_i^{k,l} = Z'_{i,k,l}\beta^{k,l} + \varepsilon_{i,k,l} \quad (5)$$

where $Z_{i,k,l}$ is a vector containing a constant term and intervention variables for child i described below. The tests for exchangeability are based on the estimates of $\beta^{k,l}$ to see whether the coefficients of the learning patterns are the same as the coefficients of other patterns.

We start by presenting results for patterns with three tasks as an example, and then report the results of all numbers of tasks, at all levels (with controls) later. Table 4 illustrates the structure of our tests for patterns of three tasks (without controls). In Table 4, we compare whether the probabilities of learning are random or not. We conduct the test conditional on the number of total tasks the child took at a given difficulty level. The tests are ordered by the number of tasks passed within the level and the number of tasks.

For example, in the first row in Table 4, at difficulty level 2, only 1 task is passed among three possible tasks. In this group, the learning pattern as we defined above is 001, and the random patterns are 010 and 100. The second row is the test for level 2, with 2 tasks passed among all the three tasks. In the data, the probability of children showing the learning pattern "011" is 0.692, and the probabilities of random patterns for "101" and "110" are 0.231 and 0.077, respectively. The null hypothesis tests $\Pr(011) = \Pr(101) = \Pr(110)$. From the Wald tests, we can reject the null at 0.01 level. Similarly, Table 4 shows the test results for other difficulty levels for three tasks.

Table 4: 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

Notes

1. “Learning patterns” are patterns with consecutive zeroes followed by consecutive ones. “Random patterns” are all other patterns.

2. The null hypotheses are that the probability of the learning patterns is equal to the probability of the random patterns. p-values < 0.1 are bolded, rejecting the null hypothesis.

3. The Chi-squared statistic and p-value are for the joint Wald test of the null hypotheses with degrees of freedom ”df”.

Next, we report the test results for patterns of all numbers of tasks at all levels with controls[11]. We control for age because maturation is a natural alternative explanation of learning. Table 5 summarizes the test results for all types of skills across all levels.[12] In Table 5 Language skill, there are altogether 72 tests, among which 71 (98.6%) reject the null hypothesis. Similarly, we can reject 60 (77.9%) out of 77 tests for Cognitive skill; 32 (84.2%) out of 38 tests for fine motor skill, and 5 (83.3%) out of 6 models for gross motor skill. Overall, we find that for most of the tests, we can **reject** the null hypothesis that the probability of random patterns equals that of the learning patterns. The empirical results are consistent for all four types of skills across all difficulty levels.

[11]Control variables include the interaction quality between home visitor and caregiver, the interaction quality between home visitor and the child, the teaching ability of home visitor, the grandmother rearing behavior and the child's age.

[12]In Appendix, Tables 8-11 contain the results for the estimates with control variables for all types of skills and all levels.

Table 5: 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%

This table shows the percentages of tests rejected by level and skill.
N/A refers to the levels we don't have enough data for.

4 Learning Pattern Features (Heterogeneity and State Dependence)

From the previous analysis, we find that there exists skill growth using children's task performance data during the intervention. In this section, we further study the properties of the children's skill growth and evaluate which of many possible learning models better fits the task performance patterns in our data.

First, we present six static models and then examine which model fits the data best. In these models, we try to add different properties (e.g., individual heterogeneity, state dependence, ability approximate measures) and then study which properties could better explain children's task performance patterns.

To compare the model fit, we first generate the predicted number of observations for each task performance pattern based on the model estimates. We then evaluate which model has a better fit with the data by comparing values of χ^2 test that compares predicted probabilities of patterns with sample probabilities.

4.1 Model Descriptions

Model 1 (Probit Model)

This model does not allow for learning. For each task at a given difficulty level, the latent process is as follows:

$$Y_i^*(t) = X'\beta + \varepsilon_{it} \tag{6}$$

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 the tasks so there is no persistent heterogeneity of ability. If $Y_i(t)$ takes the value of zero when the child cannot pass the task t

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

X consists of the same variables used in Section 3.

Model 2 (Heterogeneity)

Model 2 introduces an individual effect that persists over trials and does not allow for learning property. That is, for each task t at a given difficulty level, we have the following:

$$Y_i^*(t) = X'\beta + \theta_i + \varepsilon_{it} \tag{7}$$

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

where θ_i is the individual specific latent factor which has mean zero and variance σ_θ^2.

Model 3 (State Dependence: Learning)

Model 3 is a model of true state dependence, which captures learning (see e.g., a model influenced by Bush and Mosteller, 1955 and formally developed by Heckman, 1978, 1981) and can be represented as follows:

$$Y_i^*(t) = X'\beta + \delta \sum_{k=1}^{t-1} Y_i(k) + \varepsilon_{it} \tag{8}$$

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

where ε_{it} is i.i.d. with mean zero, and the latent value $Y_i(t)^*$ depends on the past task performance $\{Y_i(k)\}_{k=1}^{t-1}$. We use $\sum_{k=1}^{t-1} Y_i(k)$ as a measure of the past tasks' performance. This is one way to capture the notion that success produces success (a version of Bush-Mosteller reinforcement learning developed in Heckman (1978, 1981)).

Model 4 (Heterogeneity and State Dependence)

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

$$Y_i^*(t) = X'\beta + \delta \sum_{k=1}^{t-1} Y_i(k) + \theta_i + \varepsilon_{it} \tag{9}$$

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

As described previously, ε_{it} is i.i.d. with mean zero, and the latent value $Y_i(t)^*$ depends on the cumulative past task performance $\{Y_i(k)\}_{k=1}^{t-1}$ and the individual heterogeneity.

Model 5 (State Dependence and Previous Level Duration)

Model 5 is a true state dependence model that adds a time to mastery measure from previous difficulty levels as a proxy for ability. We use both the state dependence and the time to mastery measure to capture the learning property. This model can be described as:

$$Y_{i,\ell}^*(t) = X'\beta + \delta \sum_{k=1}^{t-1} Y_{i,l}(k) + \gamma D_{i,\ell-1} + \varepsilon_{it} \tag{10}$$

where ε_{it} is i.i.d. with mean zero, and the latent value $Y_{i,l}(t)^*$ at difficulty level ℓ depends on the past task performance at the same level $\{Y_{i,\ell}(k)\}_{k=1}^{t-1}$. $D_{i,\ell-1}$ represents the number of tasks required to get the first correct answer at the previous difficulty level $\ell - 1$. It is a measure of ability.

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

Model 6 (Current and Lagged State Dependence)

Model 6 is a model of true state dependence with individual unobserved heterogeneity. The difference between Model 4 and 6 is that Model 6 also includes the product term which reflects state dependence (i.e. $\delta_2 \sum_{j=1}^{t-1} \Pi_{m=1}^{j} Y_{i,\ell}(t-m)$). This is an indicator of the number of correct answers to the point. It acts like a renewal process indicator (length of current streak of successful answers) This model may be written as:

$$Y_{i,\ell}^*(t) = X'\beta + \delta \sum_{k=1}^{t-1} Y_{i,\ell}(k) + \delta_2 \sum_{j=1}^{t-1} \Pi_{m=1}^{j} Y_{i,\ell}(t-m) + \theta_i + \varepsilon_{it} \tag{11}$$

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

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

4.2 Comparing Different Model Fit

In this section, we report estimates of the six models described above. For each model, we present the results using a Wald Test.

We include all the following control variables when estimating all models: grandmother's presence, the home visitor's teaching ability, interaction quality between the home visitor and the caregiver, interaction quality between the home visitor and the child, and age.

To illustrate how to examine the model fit, in Table 6, we use fine motor skill's three tasks case from difficulty level one as an example. The column "observation" presents numbers of observations for each of the three tasks patterns from our data (fine motor skill difficulty level one). For example, "001" means the child failed the first two tasks and passed the third task at fine motor skill difficulty level one. We also report the predicted numbers of task performance patterns based on different models' estimates. We calculate the χ^2 value of each model. The row "p-value" reports the probability of the null hypothesis: the frequencies predicted by the model fit the frequencies in the data. In Tables 12-15, we report the results of χ^2 tests which examine which model's prediction fits the actual frequency of each task performance pattern the best.

Table 12 evaluates whether the models fit the actual language task patterns frequencies or not. We find that Models 4 (heterogeneity and state dependence I) and 6 (heterogeneity and state dependence II) have the best fit for task performances. This means that the models with both the heterogeneity and state dependence properties fit the data the best. (i.e., 75% of tests for Model 4 (heterogeneity and state dependence I) and 71% of tests for Model 6 (heterogeneity and state dependence II) cannot be rejected under the null hypothesis that the predicted frequency based on the model is the same as the observed frequency at 0.05 level.[13]

Table 13 reports the same analysis but for the cognitive skill tasks. We find that Models 4 and 6 also have the best performance: for Model 4, 71% of tests cannot be rejected under the null at 0.05 level, and for Model 6, 74% of tests cannot be rejected at 0.05 level.[14]. In Tables 14-15, for fine motor and gross motor skills, we also find that Model 4 has the best fit among all the models. Surprisingly, adding a proxy for ability (time to first correct answer) worsens model fit.

[13]43 tests among 57 tests across all difficulty levels for Model 4 (heterogeneity and state dependence I), and 39 tests among 55 tests for Model 6 (heterogeneity and state dependence II) can not be rejected at 0.05 level.)

[14]For Model 4, 35 tests among 49 tests cannot be rejected under the null at 0.05 level. And for Model 6, 28 among 38 tests cannot be rejected under the null at 0.05 level

We summarize the test results in Table 7. We find that Model 4 and 6 fit the date equally well. Model 4 is simpler than Model 6. Therefore, to fit the data well, a model needs to have both individual heterogeneity and state dependence properties at the same time. State dependence is a crucial pattern for language skill, and both heterogeneity and state dependence play equal roles for cognitive skill tasks. This finding also indicates that the child learning pattern during the China REACH intervention has these two properties.

Table 6: Fine Motor Skill Level 1 Cases with 3 Tasks χ^2 Test

		Predicted Number				
Pattern	Observation	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect+ State DependenceI	Model 6 Random Effect+ State DependenceII
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

Model 5 is not applicable to level 1 by design.

Table 7: 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 DependenceI	Model 5 Duration+ State Dependence	Model 6 Random Effect+ State DependenceII
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%

The values in the table are the percentages of tests which cannot reject the null hypothesis (i.e., p-value>0.05), for each model by skill.

5 Conclusion

Based on the China REACH program's exceptionally rich data, this paper examines the growth of children multiple skill development. We conduct the test of learning patterns based on the children's task performance in the intervention. Using a method of Heckman (1978), we fit linear probability models to compare the probability of child learning and find that the program significantly improves the children's multiple skills' development. Then we study how the children learn these skills by estimating an extension of the logic of Bush and Mosteller (1955) due to Heckman (1978, 1981). We find that a model with cumulative learning and child heterogeneity in latent ability fits the data best.

6 Appendix

Table 8: Hypothesis Tests for Patterns of All Tasks and All levels (Language)

Level	No.Tasks	No.Pass	Null Hypotheses	Chi-squared	p-value	D.F.	Obs.
2	3	2	Pr(011)=Pr(101)=Pr(110)	202.356	**0.000**	14	12
3	3	1	Pr(001)=Pr(100)	27.248	**0.000**	7	19
3	3	2	Pr(011)=Pr(101)=Pr(110)	59.705	**0.000**	14	20
3	4	1	Pr(0001)=Pr(0010)=Pr(1000)	19.378	**0.080**	12	11
3	4	2	Pr(0011)=Pr(0101)=Pr(0110)=Pr(1001)	62.108	**0.000**	21	15
3	4	3	Pr(0111)=Pr(1011)=Pr(1101)=Pr(1110)	104.972	**0.000**	21	14
4	3	1	Pr(001)=Pr(010)=Pr(100)	45.053	**0.000**	14	14
4	3	2	Pr(011)=Pr(101)=Pr(110)	35.366	**0.001**	14	14
4	4	1	Pr(0001)=Pr(0100)=Pr(1000)	130.084	**0.000**	14	12
4	4	2	Pr(0011)=Pr(1001)=Pr(1010)=Pr(1100)	58.075	**0.000**	21	15
4	4	3	Pr(0111)=Pr(1011)=Pr(1101)	29.229	**0.010**	14	15
4	5	2	Pr(00011)=Pr(00101)=Pr(01001)=Pr(01010)=...	99.189	**0.000**	28	15
4	5	3	Pr(00111)=Pr(01011)=Pr(01101)=Pr(01110)	119.412	**0.000**	21	10
4	6	2	Pr(000011)=Pr(000101)=Pr(001001)=...	203.607	**0.000**	28	9
5	3	1	Pr(001)=Pr(010)=Pr(100)	130.843	**0.000**	14	11
5	4	2	Pr(0011)=Pr(0101)=Pr(1010)	75685.219	**0.000**	14	9
6	3	1	Pr(001)=Pr(010)=Pr(100)	38.772	**0.000**	14	21
6	3	2	Pr(011)=Pr(101)=Pr(110)	37.641	**0.001**	14	50
6	4	1	Pr(0001)=Pr(0010)=Pr(0100)	65.826	**0.000**	14	9
6	5	2	Pr(00011)=Pr(00101)=Pr(00110)=Pr(01001)=...	390.827	**0.000**	28	10
6	6	2	Pr(000011)=Pr(000101)=Pr(001100)=...	310.003	**0.000**	35	11
6	6	3	Pr(000111)=Pr(001011)=Pr(001101)=...	94.182	**0.000**	49	17
6	6	4	Pr(001111)=Pr(011011)=Pr(101011)=...	178.594	**0.000**	56	12
6	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	68.180	**0.000**	28	20
6	7	2	Pr(0000011)=Pr(0000101)=Pr(0010001)=...	416.561	**0.000**	21	9
6	7	3	Pr(0000111)=Pr(0010011)=Pr(0100011)=...	519.619	**0.000**	35	9
6	8	5	Pr(00011111)=Pr(00101111)=Pr(00110111)=...	244.272	**0.000**	63	12
6	8	6	Pr(00111111)=Pr(01011111)=Pr(01101111)=...	796.939	**0.000**	35	10
7	8	6	Pr(00111111)=Pr(01111101)=Pr(10011111)=...	114.818	**0.000**	49	12
7	8	7	Pr(01111111)=Pr(10111111)=Pr(11111101)=...	187.311	**0.000**	21	12
7	10	8	Pr(0011111111)=Pr(0101111111)=...	289.385	**0.000**	70	12
7	10	9	Pr(0111111111)=Pr(1011111111)=...	109.632	**0.000**	42	11
8	4	2	Pr(0011)=Pr(0110)=Pr(1010)	930.595	**0.000**	14	9
8	5	3	Pr(00111)=Pr(01011)=Pr(01101)=Pr(11001)	600.386	**0.000**	21	10

Table 8: Hypothesis Tests for Patterns of All Tasks and All levels (Language)

Level	No.Tasks	No.Pass	Null Hypotheses	Chi-squared	p-value	D.F.	Obs.
8	5	4	Pr(01111)=Pr(10111)=Pr(11011)=Pr(11101)=...	158.406	**0.000**	28	12
8	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	1206.643	**0.000**	28	9
8	7	6	Pr(0111111)=Pr(1011111)=Pr(1101111)=...	96.792	**0.000**	28	12
8	8	5	Pr(00011111)=Pr(00111101)=Pr(00111110)=...	298.392	**0.000**	35	12
8	8	6	Pr(00111111)=Pr(01011111)=Pr(01110111)=...	109.636	**0.000**	63	18
8	9	6	Pr(000111111)=Pr(001011111)=...	155.160	**0.000**	56	14
8	9	7	Pr(001111111)=Pr(010111111)=...	194.978	**0.000**	77	13
8	9	8	Pr(011111111)=Pr(101111111)=...	95.173	**0.000**	28	10
8	10	7	Pr(0001111111)=Pr(0011111011)=...	301.031	**0.000**	42	10
8	10	8	Pr(0011111111)=Pr(0101111111)=...	375.124	**0.000**	63	15
8	10	9	Pr(0111111111)=Pr(1011111111)=...	116.437	**0.000**	42	16
8	11	7	Pr(00001111111)=Pr(00011111101)=...	364.849	**0.000**	70	13
8	11	9	Pr(00111111111)=Pr(01111101111)=...	610.234	**0.000**	84	15
8	11	10	Pr(01111111111)=Pr(10111111111)=...	863.789	**0.000**	28	10
9	3	1	Pr(001)=Pr(010)=Pr(100)	47.952	**0.000**	14	10
9	3	2	Pr(011)=Pr(101)=Pr(110)	21.016	0.101	14	19
9	4	2	Pr(0011)=Pr(0101)=Pr(0110)=Pr(1001)=Pr(1100)	82.238	**0.000**	28	15
9	4	3	Pr(0111)=Pr(1011)=Pr(1101)=Pr(1110)	60.647	**0.000**	21	22
9	5	1	Pr(00001)=Pr(00100)=Pr(01000)=Pr(10000)	465.852	**0.000**	21	11
9	5	3	Pr(00111)=Pr(01101)=Pr(01110)=Pr(10011)=...	2037.542	**0.000**	28	9
9	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	59.063	**0.007**	35	14
9	7	5	Pr(0011111)=Pr(0101111)=Pr(0111110)=...	107.959	**0.000**	35	14
9	7	6	Pr(0111111)=Pr(1011111)=Pr(1101111)=...	330.868	**0.000**	28	12
10	3	2	Pr(011)=Pr(101)=Pr(110)	54.468	**0.000**	12	10
10	4	3	Pr(0111)=Pr(1011)=Pr(1101)=Pr(1110)	142.476	**0.000**	21	9
10	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	170.201	**0.000**	21	10
10	7	5	Pr(0011111)=Pr(0101111)=Pr(0110111)=...	315.256	**0.000**	49	15
10	7	6	Pr(1011111)=Pr(1101111)=Pr(1110111)=...	267.277	**0.000**	21	10
10	8	6	Pr(00111111)=Pr(10011111)=Pr(10101111)=...	1987.754	**0.000**	42	15
11	3	1	Pr(001)=Pr(010)=Pr(100)	75.648	**0.000**	14	27
11	3	2	Pr(011)=Pr(101)=Pr(110)	39.345	**0.000**	14	42
11	5	1	Pr(00001)=Pr(00010)=Pr(00100)=Pr(10000)	530.381	**0.000**	21	11
11	5	4	Pr(01111)=Pr(10111)=Pr(11011)=Pr(11101)=...	67.846	**0.000**	28	21
11	6	4	Pr(001111)=Pr(010111)=Pr(011011)=...	278.696	**0.000**	56	13
11	6	5	Pr(011111)=Pr(101111)=Pr(111011)=...	98.357	**0.000**	28	25

Table 8: Hypothesis Tests for Patterns of All Tasks and All levels (Language)

Level	No.Tasks	No.Pass	Null Hypotheses	Chi-squared	p-value	D.F.	Obs.
11	7	4	Pr(0001111)=Pr(0011110)=Pr(0101101)=...	1658.925	**0.000**	42	10
11	7	5	Pr(0011111)=Pr(0101111)=Pr(0110111)=...	260.512	**0.000**	70	18
11	7	6	Pr(0111111)=Pr(1011111)=Pr(1101111)=...	87.315	**0.000**	42	23

Note: 1. The "No. Tasks" column gives the length of the pattern.
2. The "No. Pass" column gives the number of tasks passed in each of the patterns in the null hypotheses.
3. "Learning patterns" are patterns with consecutive zeroes followed by consecutive ones.
"Random patterns" are all other patterns.
4. The null hypotheses are that the probability of the learning patterns is equal to the probability of the random patterns. p-values < 0.1 are bolded, rejecting the null hypothesis.
5. The Chi-squared statistic and p-value are for the joint Wald test of the null hypotheses with degrees of freedom "df".

Table 9: Hypothesis Tests for Patterns of All Tasks and All levels (Cognitive)

Level	No.Task	No.Pass	Hypothesis	Chi-squared	p-value	D.F.	Obs.
2	3	1	Pr(001)=Pr(010)	5.706	0.574	7	20
2	3	2	Pr(011)=Pr(101)	25.184	**0.001**	7	14
2	6	1	Pr(000001)=Pr(000010)=Pr(000100)=...	47.395	**0.012**	28	14
2	6	3	Pr(000111)=Pr(001110)=Pr(010011)=...	45.944	0.598	49	13
2	6	4	Pr(001111)=Pr(011011)=Pr(101011)=...	32.665	0.248	28	9
2	6	5	Pr(011111)=Pr(101111)=Pr(110111)	62.129	**0.000**	14	12
2	7	1	Pr(0000001)=Pr(0000010)=Pr(0000100)=...	32.344	**0.054**	21	9
2	7	2	Pr(0000011)=Pr(0000101)=Pr(0010010)=...	41.335	0.500	42	11
2	7	3	Pr(0000111)=Pr(0011100)=Pr(0100011)=...	140.186	**0.000**	35	14
2	8	1	Pr(00000001)=Pr(00000010)=Pr(00000100)=...	42.448	**0.004**	21	9
2	8	2	Pr(00000011)=Pr(00000110)=Pr(00001001)=...	440.775	**0.000**	70	19
2	8	4	Pr(00001111)=Pr(00010111)=Pr(00100111)=...	1281.095	**0.000**	77	22
2	8	6	Pr(00111111)=Pr(01011111)=Pr(10101111)=...	55.343	0.500	56	15
2	8	7	Pr(01111111)=Pr(10111111)=Pr(11101111)=...	82.568	**0.000**	21	12
5	6	5	Pr(011111)=Pr(110111)=Pr(111101)	42.209	**0.000**	14	12
5	7	5	Pr(0011111)=Pr(0101111)=Pr(0111011)=...	60.836	0.306	56	12
5	8	2	Pr(00000011)=Pr(00001001)=Pr(00100001)=...	44.980	**0.002**	21	9
5	8	7	Pr(01111111)=Pr(10111111)=Pr(11011111)=...	115.562	**0.000**	21	15
5	9	1	Pr(000000001)=Pr(000000010)=...	55.310	**0.002**	28	9
5	9	6	Pr(000111111)=Pr(010011111)=...	84.807	0.826	98	15

Table 9: Hypothesis Tests for Patterns of All Tasks and All levels (Cognitive)

Level No.	Task No.	Pass	Hypothesis	Chi-squared	p-value	D.F.	Obs.
5	9	8	Pr(011111111)=Pr(111011111)=...	93.991	**0.000**	35	20
5	10	1	Pr(0000000001)=Pr(0000000100)=...	20.730	0.109	14	9
5	10	4	Pr(0000001111)=Pr(0000100111)=...	56.056	0.720	63	11
5	10	7	Pr(0001111111)=Pr(0011011111)=...	84.689	0.458	84	14
5	10	9	Pr(0111111111)=Pr(1110111111)=...	29.813	**0.096**	21	23
6	3	2	Pr(011)=Pr(101)	34.045	**0.000**	7	12
6	4	3	Pr(0111)=Pr(1011)	9.539	0.216	7	19
6	5	2	Pr(00011)=Pr(01001)	70.332	**0.000**	7	10
6	5	3	Pr(00111)=Pr(01011)=Pr(01101)	66.672	**0.000**	14	12
6	5	4	Pr(01111)=Pr(10111)=Pr(11011)=Pr(11101)	47.306	**0.001**	21	28
6	6	2	Pr(000011)=Pr(000101)=Pr(000110)	36.203	**0.001**	14	9
6	6	3	Pr(000111)=Pr(001011)=Pr(001101)=...	93.652	**0.000**	49	28
6	6	4	Pr(001111)=Pr(010111)=Pr(011011)=...	149.402	**0.000**	56	18
6	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	38.373	**0.092**	28	37
6	7	3	Pr(0000111)=Pr(0001011)=Pr(0001101)=...	58.980	**0.001**	28	22
6	7	4	Pr(0001111)=Pr(0010111)=Pr(0011011)=...	230.892	**0.000**	42	22
6	7	5	Pr(0011111)=Pr(0101111)=Pr(0110111)=...	635.021	**0.000**	70	28
6	7	6	Pr(0111111)=Pr(1011111)=Pr(1101111)=...	58.837	**0.007**	35	47
7	7	6	Pr(0111111)=Pr(1011111)=Pr(1110111)	43.747	**0.000**	14	13
7	8	7	Pr(01111111)=Pr(10111111)=Pr(11011111)=...	889.317	**0.000**	21	11
7	9	6	Pr(000111111)=Pr(001011111)=...	65.062	0.645	70	11
7	9	7	Pr(001111111)=Pr(010111111)=...	71.810	**0.076**	56	11
7	9	8	Pr(011111111)=Pr(101111111)=...	172.742	**0.000**	49	27
7	10	7	Pr(0001111111)=Pr(0010111111)=...	102.704	**0.007**	70	16
7	10	8	Pr(0011111111)=Pr(0101111111)=...	825.801	**0.000**	98	25
7	10	9	Pr(0111111111)=Pr(1011111111)=...	165.188	**0.000**	49	24
7	11	7	Pr(00001111111)=Pr(00010111111)=...	105.481	**0.004**	70	17
7	11	9	Pr(00111111111)=Pr(01101111111)=...	87.134	**0.000**	42	13
7	11	10	Pr(01111111111)=Pr(10111111111)=...	103.114	**0.000**	49	19
8	3	2	Pr(011)=Pr(101)	22.008	**0.003**	7	12
8	4	2	Pr(0011)=Pr(0101)=Pr(1001)=Pr(1010)	78.244	**0.000**	21	17
8	4	3	Pr(0111)=Pr(1011)=Pr(1101)	41.903	**0.000**	14	15
8	5	1	Pr(00001)=Pr(00100)=Pr(01000)	34.494	**0.002**	14	9
8	5	2	Pr(00110)=Pr(01100)=Pr(10001)=Pr(10100)	202.828	**0.000**	21	11
8	5	4	Pr(01111)=Pr(10111)=Pr(11011)=Pr(11101)	74.564	**0.000**	21	32

Table 9: Hypothesis Tests for Patterns of All Tasks and All levels (Cognitive)

Level	No.Task	No.Pass	Hypothesis	Chi-squared	p-value	D.F.	Obs.
8	6	2	Pr(000011)=Pr(000101)=Pr(000110)=...	397.755	**0.000**	49	16
8	6	3	Pr(000111)=Pr(001011)=Pr(001101)=...	344.548	**0.000**	56	18
8	6	4	Pr(001111)=Pr(011011)=Pr(011101)=...	75.724	**0.041**	56	15
8	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	51.882	**0.004**	28	26
8	7	3	Pr(0000111)=Pr(0001011)=Pr(0001101)=...	70.109	**0.004**	42	13
8	7	4	Pr(0001111)=Pr(0011101)=Pr(0011110)=...	50.827	0.401	49	11
8	7	5	Pr(0011111)=Pr(0101111)=Pr(0111011)=...	446.094	**0.000**	77	42
8	7	6	Pr(0111111)=Pr(1011111)=Pr(1101111)=...	84.432	**0.000**	35	38
10	3	2	Pr(011)=Pr(101)	16.558	**0.020**	7	20
10	4	2	Pr(0011)=Pr(0101)=Pr(1001)	62.007	**0.000**	14	22
10	4	3	Pr(0111)=Pr(1011)	11.777	0.108	7	16
11	3	1	Pr(001)=Pr(010)	11.156	0.132	7	14
11	4	1	Pr(0001)=Pr(0010)=Pr(0100)	81.847	**0.000**	14	11
11	4	3	Pr(0111)=Pr(1011)=Pr(1101)	47.386	**0.000**	14	13
11	5	2	Pr(00011)=Pr(00101)=Pr(01001)=Pr(01010)=...	177.726	**0.000**	35	17
11	5	4	Pr(01111)=Pr(10111)=Pr(11011)=Pr(11101)	688.177	**0.000**	21	12
11	6	2	Pr(000110)=Pr(001001)	36.764	**0.000**	6	9
11	6	3	Pr(000111)=Pr(001011)=Pr(011010)	545.873	**0.000**	14	14
11	6	4	Pr(001111)=Pr(101011)=Pr(101101)=...	67.512	**0.001**	35	12
11	6	5	Pr(011111)=Pr(101111)=Pr(111011)=...	25.047	0.245	21	18
12	3	1	Pr(001)=Pr(010)	8.583	0.284	7	16
12	3	2	Pr(011)=Pr(101)	12.169	**0.095**	7	33

Note: 1. The "No. Tasks" column gives the length of the pattern.
2. The "No. Pass" column gives the number of tasks passed in each of the patterns in the null hypotheses.
3. "Learning patterns" are patterns with consecutive zeroes followed by consecutive ones.
"Random patterns" are all other patterns.
4. The null hypotheses are that the probability of the learning patterns is equal to the probability of the random patterns. p-values < 0.1 are bolded, rejecting the null hypothesis.
5. The Chi-squared statistic and p-value are for the joint Wald test of the null hypotheses with degrees of freedom "df".

Table 10: Hypothesis Tests for Patterns of All Tasks and All levels (Fine Motor)

Level	No.Tasks	No.Pass	Null Hypotheses	Chi-squared	p-value	D.F.	Obs.
1	3	1	Pr(001)=Pr(010)	11.959	0.102	7	14
1	3	2	Pr(011)=Pr(101)	8.425	0.297	7	18
1	4	2	Pr(0011)=Pr(0101)=Pr(0110)=Pr(1001)	143.650	**0.000**	21	12
1	4	3	Pr(0111)=Pr(1011)	58.728	**0.000**	7	10
1	5	2	Pr(00011)=Pr(00101)=Pr(00110)=Pr(01001)=...	48.650	**0.062**	35	9
1	5	3	Pr(00111)=Pr(01101)=Pr(01110)=Pr(10011)	99.250	**0.000**	21	12
1	5	4	Pr(01111)=Pr(10111)=Pr(11011)=Pr(11101)	390.653	**0.000**	21	11
1	6	3	Pr(000111)=Pr(001101)=Pr(010101)=...	56.550	**0.000**	21	10
1	6	4	Pr(001111)=Pr(010111)=Pr(011101)=...	80.060	**0.000**	42	16
1	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	56.145	**0.001**	28	19
3	3	2	Pr(011)=Pr(101)	23.977	**0.001**	7	15
3	4	2	Pr(0011)=Pr(0101)=Pr(0110)	27.914	**0.015**	14	11
3	4	3	Pr(0111)=Pr(1011)=Pr(1101)	50.377	**0.000**	14	25
3	5	1	Pr(00001)=Pr(00100)=Pr(01000)	32.896	**0.003**	14	14
3	5	2	Pr(00011)=Pr(00101)=Pr(00110)=Pr(01001)=...	87.269	**0.000**	42	20
3	5	3	Pr(00111)=Pr(01011)=Pr(01101)=Pr(01110)=...	262.951	**0.000**	42	23
3	5	4	Pr(01111)=Pr(10111)=Pr(11011)=Pr(11101)	74.213	**0.000**	21	54
3	6	1	Pr(000001)=Pr(000100)=Pr(001000)=...	102.917	**0.000**	21	15
3	6	2	Pr(000011)=Pr(000110)=Pr(001001)=...	379.227	**0.000**	42	14
3	6	3	Pr(000111)=Pr(001011)=Pr(001101)=...	23266.533	**0.000**	77	24
3	6	4	Pr(001111)=Pr(010111)=Pr(011011)=...	175.314	**0.000**	77	50
3	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	58.808	**0.001**	28	45
4	3	2	Pr(011)=Pr(101)	9.228	0.237	7	11
4	4	3	Pr(0111)=Pr(1011)=Pr(1101)	29.114	**0.010**	14	21
4	5	2	Pr(00011)=Pr(00101)=Pr(00110)=Pr(01100)=...	52.293	**0.030**	35	10
4	5	3	Pr(00111)=Pr(01011)=Pr(01101)=Pr(01110)=...	48.490	**0.009**	28	10
4	5	4	Pr(01111)=Pr(11011)=Pr(11101)	136.655	**0.000**	14	19
4	6	2	Pr(000011)=Pr(000101)=Pr(001010)=...	35.332	**0.026**	21	9
4	6	4	Pr(001111)=Pr(010111)=Pr(011011)=...	2443.406	**0.000**	70	22
4	6	5	Pr(011111)=Pr(101111)=Pr(110111)=...	30.762	0.328	28	47
5	3	1	Pr(001)=Pr(010)	9.924	0.193	7	15
5	3	2	Pr(011)=Pr(101)	12.678	**0.080**	7	41
6	3	1	Pr(001)=Pr(010)	90.577	**0.000**	7	22
6	3	2	Pr(011)=Pr(101)	20.961	**0.004**	7	13
6	4	2	Pr(0011)=Pr(0101)=Pr(0110)=Pr(1001)	36.042	**0.022**	21	14

Table 10: Hypothesis Tests for Patterns of All Tasks and All levels (Fine Motor)

Level	No.Tasks	No.Pass	Null Hypotheses	Chi-squared	p-value	D.F.	Obs.
6	4	3	Pr(0111)=Pr(1011)=Pr(1101)	44.176	**0.000**	14	25
7	3	1	Pr(001)=Pr(010)	57.569	**0.000**	7	24
7	3	2	Pr(011)=Pr(101)	4.877	0.675	7	24

Note: 1. The "No. Tasks" column gives the length of the pattern.
2. The "No. Pass" column gives the number of tasks passed in each of the patterns in the null hypotheses.
3. "Learning patterns" are patterns with consecutive zeroes followed by consecutive ones.
"Random patterns" are all other patterns.
4. The null hypotheses are that the probability of the learning patterns is equal to the probability of the random patterns. p-values < 0.1 are bolded, rejecting the null hypothesis.
5. The Chi-squared statistic and p-value are for the joint Wald test of the null hypotheses with degrees of freedom "df".

Table 11: Hypothesis Tests for Patterns of All Tasks and All levels (Gross Motor)

Level	No.Tasks	No.Pass	Null Hypotheses	Chi-squared	p-value	D.F.	Obs.
4	3	1	Pr(001)=Pr(010)	21.645	**0.003**	5	25
4	3	2	Pr(011)=Pr(101)	48.779	**0.000**	5	33
7	3	1	Pr(001)=Pr(010)	13.709	**0.057**	5	23
7	3	2	Pr(011)=Pr(101)	9.382	0.226	5	52
8	3	1	Pr(001)=Pr(010)	14.705	**0.040**	5	62
8	3	2	Pr(011)=Pr(101)	22.471	**0.002**	5	86

Note: 1. The "No. Tasks" column gives the length of the pattern.
2. The "No. Pass" column gives the number of tasks passed in each of the patterns in the null hypotheses.
3. "Learning patterns" are patterns with consecutive zeroes followed by consecutive ones.
"Random patterns" are all other patterns.
4. The null hypotheses are that the probability of the learning patterns is equal to the probability of the random patterns. p-values < 0.1 are bolded, rejecting the null hypothesis.
5. The Chi-squared statistic and p-value are for the joint Wald test of the null hypotheses with degrees of freedom "df".

Table 12: Chi-Squared Test p-Values for Task Performance Patterns, All Language Levels (with All Controls)

Level	No. Tasks	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect+ State DependenceI	Model 5 Duration+ State Dependence	Model 6 Random Effect+ State DependenceII
2	3	0.102	0.103	0.645	0.645		0.573
	4	**0.003**	0.725	0.199	0.797		0.767
	5	**0.000**	0.068	0.216	0.216		0.178
	6	**0.038**	0.475	0.945	0.936		0.936
	7	0.451	0.758	0.604	0.597		0.772
	8	**0.012**	0.406	1.000			1.000
	9	**0.000**	0.848	1.000	1.000		1.000
	10	**0.000**	**0.014**	**0.022**	0.518		0.054
3	3	**0.000**	**0.001**	**0.017**	**0.024**	**0.000**	**0.026**
	4	**0.000**	**0.001**	0.057	0.058	**0.006**	**0.032**
	5	**0.000**	0.138	0.604	0.608	**0.017**	0.639
4	3	0.415	0.737	0.841	0.852	**0.000**	0.797
	4	**0.006**	**0.018**	**0.030**	**0.026**	**0.000**	0.083
	5	**0.000**	**0.001**	0.130	0.130	**0.000**	0.111
	6	**0.000**	**0.020**	0.514	0.514	0.219	0.405
	7	**0.000**		0.050		**0.000**	
5	3	0.164	0.499	0.452	0.477	0.266	0.550
	4	**0.000**	**0.000**	**0.023**	**0.022**	**0.001**	**0.000**
	5	0.779	0.794	0.818	0.812	0.452	0.699
	6	**0.000**	**0.012**	0.133	0.174	0.131	0.132
6	3	**0.000**	**0.010**	**0.000**	**0.000**		**0.000**
	4	**0.023**	**0.004**	0.487	0.487		
	5	**0.000**	**0.000**	**0.000**	**0.000**		**0.000**
	6	**0.000**	**0.000**	**0.040**	0.658	**0.000**	0.631
	7	1.000	**0.000**	1.000	1.000	**0.000**	1.000
	8	**0.000**	**0.002**	1.000	1.000	**0.001**	1.000
	9	1.000	**0.000**	1.000	1.000	0.790	1.000
7	3	0.422		0.789	0.764		0.480
	4	0.932	0.892	0.987	0.987		0.983
	5	0.050	**0.032**	0.337	0.336		
	6	0.799	0.801	0.624	0.611	**0.011**	0.400
	7	0.289	0.659	0.530	0.743	**0.000**	0.675
	8	**0.000**	**0.008**	**0.000**	0.197	**0.000**	0.204
	9	1.000	0.615	1.000	1.000	0.252	1.000
	10	0.989	**0.000**	0.999	1.000	**0.000**	1.000

Table 12: Chi-Squared Test p-Values for Task Performance Patterns, All Language Levels (with All Controls)

Level	No. Tasks	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect+ State Dependence	Model 5 Duration+ State Dependence	Model 6 Random Effect+ State DependenceII
8	3	**0.001**	**0.007**	0.182	0.182	**0.000**	0.640
	4	**0.004**	**0.002**	0.142	0.142	0.082	**0.006**
	5	**0.000**	**0.000**	0.117	0.117	**0.043**	**0.047**
	6	**0.003**	**0.000**	**0.002**	**0.002**	**0.015**	**0.000**
	7	0.324	0.148	**0.005**	**0.021**	**0.000**	**0.038**
	8	**0.000**	**0.000**	0.512	0.622	**0.000**	0.973
	9	1.000	**0.002**	**0.000**	0.990	**0.000**	0.999
	10	1.000	0.977	1.000	1.000	**0.000**	1.000
9	3	0.924	0.988	0.960	0.992	0.984	0.993
	4	**0.017**	**0.020**	0.351	0.372	0.332	0.395
	5	**0.008**	**0.002**	**0.001**	**0.004**	**0.001**	**0.019**
	6	**0.027**	0.368	**0.012**	0.452	**0.010**	0.524
	7	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.019**
10	3	0.220		0.428	0.413	**0.032**	0.398
	4	0.924	0.869	0.979	0.980	0.911	
	5	**0.007**	0.118	**0.005**	**0.006**	**0.000**	**0.006**
	6	0.054	**0.009**	0.136	0.137	**0.008**	0.119
	7	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	8	1.000	**0.000**	1.000	1.000	**0.000**	1.000
11	3	**0.001**	0.056	**0.034**	0.150	**0.000**	0.243
	4	**0.037**	0.326	0.105	0.487	0.199	0.636
	5	**0.001**	**0.019**	**0.001**	**0.019**	**0.000**	**0.016**
	6	**0.045**	**0.005**	**0.004**	**0.008**	**0.000**	**0.010**
	7	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
#p-values>0.05		22	23	39	**43**	11	**39**
Num. of Tests		59	56	59	57	45	55
Frac of test #p-val>0.05		0.373	0.411	0.661	**0.754**	0.244	**0.709**

Note: 1. p-values < 0.05 are bolded, rejecting the hypothesis that the data is generated by the model in that column.
2. By construction, Model 5 doesn't start from the 1st level in the table since they rely on the duration of the previous level.
3. By construction, Model 6 is not applicable to patterns with only 2 tasks.
4. Other missing cells are due to lack of observations.

Table 13: Chi-Squared Test p-Values for Task Performance Patterns, All Cognitive skills Levels (with All Controls)

Level	No. Tasks	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect+ State DependenceI	Model 5 Duration+ State Dependence	Model 6 Random Effect+ State DependenceII
2	2	**0.002**	0.890	**0.001**	0.942		
	3	0.317	0.471	0.343	0.501		0.524
	4	0.316	0.315	0.236	0.540		0.530
	5	0.875	0.835	0.886	0.863	**0.000**	0.891
	6	**0.037**	**0.027**	**0.001**	**0.045**	**0.000**	0.273
	7	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.008**
	8	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
3	2	**0.000**	0.887	**0.000**	0.951	**0.000**	
4	2	**0.000**	**0.002**	**0.000**	0.184	**0.000**	
5	2	**0.000**	0.487	0.584	0.509		
	3			0.482			
	4	0.286	0.953	0.081	0.978		0.974
	5	**0.000**	**0.000**	**0.000**	0.060		0.051
	6	0.255	0.221	0.325	0.233	**0.000**	0.249
	7	**0.006**	0.691	0.305	0.633	**0.000**	0.709
	8	**0.000**	**0.038**	**0.000**	**0.048**	**0.000**	0.083
	9	**0.000**	0.241	**0.000**	0.287	**0.000**	0.333
	10	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.003**
6	2	0.206	0.133	0.666	0.624	0.377	
	3	**0.003**	0.170	0.064	0.479	**0.000**	0.459
	4	**0.005**	**0.003**	0.107	0.281	0.086	0.246
	5	**0.000**	**0.000**	**0.000**	**0.001**	**0.001**	**0.002**
	6	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	7	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
7	2	**0.018**		**0.013**	**0.004**	**0.000**	
	3	**0.009**	0.824	0.245	0.936	0.753	0.918
	4	0.885	0.781	0.777	0.848	0.670	0.887
	5	0.110	**0.021**	0.297	0.307	0.290	0.149
	6	**0.000**	0.286	0.439	0.261	0.073	0.451
	7	0.062	**0.035**	0.473	0.326	0.089	0.470
	8	0.366	0.597	0.914	0.959	0.821	0.956
	9	**0.000**	**0.012**	**0.000**	0.679	**0.017**	0.678
	10	0.331	**0.000**	**0.000**	**0.033**	**0.000**	0.091
8	2	0.320	0.402	0.409	0.501	0.405	

Table 13: Chi-Squared Test p-Values for Task Performance Patterns, All Cognitive skills Levels (with All Controls)

Level	No. Tasks	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect+ State DependenceI	Model 5 Duration+ State Dependence	Model 6 Random Effect+ State DependenceII
	3	**0.044**	0.369	0.132	0.556	**0.018**	0.659
	4	**0.010**	0.142	**0.010**	0.186	0.087	0.173
	5	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**
	6	**0.000**	**0.020**	**0.009**	0.215	**0.027**	0.235
	7	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
9	2	0.935	0.957	0.944	0.872	0.497	
10	2	**0.000**	0.706	**0.000**	0.829	**0.000**	
	3	0.396	0.684	0.654	0.676	**0.000**	0.601
	4	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
11	2	0.143	0.586	0.215	0.601	**0.000**	
	3	0.057	0.169	0.129	0.214	**0.000**	0.180
	4	0.227	0.580	0.403	0.620	0.087	0.619
	5	**0.001**	**0.002**	**0.031**	0.159	**0.036**	0.185
	6	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
12	2	0.078	0.069	0.606	0.606	0.651	
	3	0.125	0.115	0.066	0.088	**0.024**	0.496
#p-values>0.05		19	27	26	**35**	13	28
Num. of Tests		49	48	50	49	43	38
Frac of test #p-val>0.05		0.388	0.563	0.52	**0.714**	0.302	**0.737**

Note: 1. p-values < 0.05 are bolded, rejecting the hypothesis that the data is generated by the model in that column.
2. By construction, Model 5 doesn't start from the 1st level in the table since they rely on the duration of the previous level.
3. By construction, Model 6 is not applicable to patterns with only 2 tasks.
4. Other missing cells are due to lack of observations.

Table 14: Chi-Squared Test p-Values for Task Performance Patterns, All Fine Motor Skill Levels (with All Controls)

Level	No. Tasks	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect+ State DependenceI	Model 5 Duration+ State Dependence	Model 6 Random Effect+ State DependenceII
1	2	**0.023**	0.094	**0.029**	0.114		
	3	0.145	0.366	0.293	0.449		0.347
	4	0.146	0.416	0.692	0.649		0.565
	5	**0.002**	0.052	0.398	0.415		0.132
	6	**0.000**	**0.000**	0.272	0.360		0.381
2	2	**0.000**	**0.028**	**0.000**	**0.011**	**0.000**	
3	2	0.401	0.384	0.840	0.838	**0.000**	
	3	0.124	0.097	0.872	0.797		0.819
	4	**0.000**	**0.001**	**0.000**	0.069	**0.000**	0.094
	5	**0.000**	**0.000**	**0.000**	**0.009**	**0.000**	**0.018**
	6	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
4	2	0.761	0.841	0.388	0.932	0.487	
	3	0.833		0.694	0.803	0.815	0.638
	4	**0.000**	**0.004**	**0.001**	**0.008**	**0.007**	**0.042**
	5	**0.000**	**0.041**	**0.000**	0.111	**0.000**	0.166
	6	**0.000**	**0.015**	**0.000**	0.068	**0.000**	0.100
5	2	**0.000**	**0.000**	0.060	**0.045**	0.228	
	3	0.384	0.402	0.711	0.735	0.525	0.672
6	2	0.154	0.148	0.636	0.637	0.179	
	3	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	4	0.077	0.153	0.747	0.644	0.102	0.640
7	2	0.183	0.189	0.993	0.821	0.802	
	3	0.056	0.066	**0.043**	0.209	**0.006**	0.230
#p-values>0.05		11	12	13	17	7	12
Num. of Tests		23	22	23	23	17	16
Frac of test #p-val>0.05		0.478	0.545	0.565	0.739	0.412	0.75

Note: 1. p-values < 0.05 are bolded, rejecting the hypothesis that the data is generated by the model in that column.
2. By construction, Model 5 doesn't start from the 1st level in the table since they rely on the duration of the previous level.
3. By construction, Model 6 is not applicable to patterns with only 2 tasks.
4. Other missing cells are due to lack of observations.

Table 15: Chi-Squared Test p-Values for Task Performance Patterns, All Gross Motor Levels (with All Controls)

Level	No. Tasks	Model 1 Probit	Model 2 Random Effect	Model 3 State Dependence	Model 4 Random Effect+ State DependenceI	Model 5 Duration+ State Dependence	Model 6 Random Effect+ State DependenceII
2	2	0.611	0.858	0.940	0.939		
	3	0.945					
4	2	0.582	0.756	0.660	0.740		
	3	**0.000**	**0.000**	0.062	0.072		0.068
5	2	**0.000**	**0.000**	**0.000**		**0.000**	
6	2	**0.000**	**0.000**	0.054	0.142	**0.000**	
7	2	**0.011**	0.193	**0.023**	0.176	**0.000**	
	3	**0.001**	0.110	**0.000**	0.115	**0.000**	0.104
8	2	0.269	0.530	0.182	0.506	0.795	
	3	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.006**
9	2	**0.000**	**0.017**	**0.000**	0.117	**0.000**	
#p-values>0.05		4	5	5	8	0	2
Num. of Tests		11	10	10	9	7	3
Frac of test #p-val>0.05		0.363	0.5	0.5	0.889	0	0.667

Note: 1. p-values <0.05 are bolded, rejecting the hypothesis that the data is generated by the model in that column.
2. By construction, Model 5 doesn't start from the 1st level in the table since they rely on the duration of the previous level.
3. By construction, Model 6 is not applicable to patterns with only 2 tasks.
4. Other missing cells are due to lack of observations.

References

Bush, Robert R. and Frederick Mosteller. 1955. *Stochastic models for learning.* John Wiley & Sons Inc. 9, 4.1, 5

Gertler, Paul, James J. Heckman, Rodrigo Pinto, Susan M. Chang-Lopez, Sally Grantham-McGregor, Christel Vermeersch, Susan Walker, and Amika S. Wright. 2021. "Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31." Unpublished. 2, 2

Gertler, Paul, James J. Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan Chang, and Sally M. Grantham-McGregor. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344 (6187):998–1001. 2

Grantham-McGregor, Sally and Joanne A. Smith. 2016. "Extending The Jamaican Early Childhood Development Intervention." *Journal of Applied Research on Children: Informing Policy for Children at Risk* 7 (2). 1, 2

Heckman, James and Jin Zhou. 2021. "Interaction as Investment: The Microdynamics and Measurement of Early Childhood Investment." *Working paper* . 2.2, 2.3

Heckman, James J. 1978. "Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence." *Annales d l'INSEE* 30–31:227–269. Special Issue. 1, 3.1, 9, 3.2, 4.1, 4.1, 5

———. 1981. "Statistical Models for Discrete Panel Data." In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. Manski and D. McFadden. Cambridge, MA: MIT Press, 114–178. 3.1, 9, 4.1, 4.1, 5

HomVEE. 2020. "Early Childhood Home Visiting Models: Reviewing Evidence of Effectiveness, 2011-2020." OPRE Report 2020-126. 1

Howard, Kimberly S. and Jeanne. Brooks-Gunn. 2009. "The Role of Home-Visiting Programs in Preventing Child Abuse and Neglect." *The Future of Children* 19 (2):119–146. 1

Palmer, Francis H. 1971. *Concept training curriculum for children ages two to five.* Stony Brook, NY: State University of New York at Stony Brook. 2.2, 2.3, 2.3

Restle, Frank and James G. Greeno. 1970. *Introduction to Mathematical Psychology.* Addison-Wesley. 9

Uzgiris, Ina C. and Joseph Hunt. 1975. "Assessment in Infancy: Ordinal Scales of Psychological Development." *University of Illinois Press* . 2.2, 2.3

Zhou, Jin, James Heckman, Bei Liu, and Mai Lu. 2021. "Treatment Effects and the Measurement of Skills in an Influential Home Visiting Program." Working Paper. 1