ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 95/25

Dynamic Complementarity

James Heckman, Haihan Tian, Zijian Zhang, Jin Zhou

www.rfberlin.com NOVEMBER 2025

Dynamic Complementarity

Authors

James Heckman, Haihan Tian, Zijian Zhang, Jin Zhou

Reference

JEL Codes: J1, D83, C1, C5

Keywords: Dynamic Complementarity, Learning, Human Capital

Recommended Citation: James Heckman, Haihan Tian, Zijian Zhang, Jin Zhou (2025): Dynamic Complementarity. RFBerlin Discussion Paper No. 95/25

Access

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Dynamic Complementarity*

James Heckman, Haihan Tian, Zijian Zhang, and Jin Zhou

October 6, 2025

*This research was supported in part by NIH grants NICHD R37HD065072 and NIH 1R01HD103666, Hong Kong RGC ECS Grant 9048312, the China Development Research Foundation, the University of Chicago's Provost Global Faculty Award, and an anonymous donor. The program data has been registered at AEA with registry number AEARCTR-0007119. We thank our partners on the China REACH project: China Development Research Foundation (CDRF), Beijing Normal University, Child Development Research Group at the University of West Indies, and Dr. Sally Grantham-McGregor. Chris Walters, the editor, and two anonymous referees made valuable comments on the first version of the paper. We thank the editor and an anonymous referee for helpful comments on the second revision of this paper. Flavio Cunha, Tamara McGavock, and Fanmei Xia made valuable suggestions. 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. James J. Heckman: Center for the Economics of Human Development, University of Chicago Department of Economics, jheckman@uchicago.edu; Haihan Tian: Center for the Economics of Human Development, University of Chicago Department of Economics, hhtian@uchicago.edu; Zijian Zhang: Columbia University, zz3263@columbia.edu; Jin Zhou: Department of Economics and Finance, City University of Hong Kong, jin.zhou@cityu.edu.hk. The web appendix can be found at http://cehd.uchicago.edu/dynamic_complementarity_appx.

Abstract

Dynamic complementarity is the concept that past investments that lead to higher stocks of skill at an age, promote the growth of skills from investment at that age. We define and produce evidence on dynamic complementarity and its three components using unique Chinese data from a home visiting program for young children targeted to parents in rural China. In addition, we investigate growth in learning due to innate, parental, and environmental factors that occur in the absence of any formal intervention.

JEL Codes: J1, D83, C1, C5
Keywords: dynamic complementarity, learning, human capital

Introduction

Dynamic complementarity is a central concept in human development. It characterizes how early learning experiences affect subsequent learning and achievement. It addresses the policy question of whether those exposed late to learning environments can catch up with those who start early.

The technology of skill formation is a framework that enables analysts to address these questions in a rigorous way. One version of it is a multistage technology in which a vector of skills at age $a+1$, $\boldsymbol{K}(a+1)$, depends on previous levels of skill and investment:

$$\boldsymbol{K}(a+1) = \boldsymbol{f}^{(a)}(\boldsymbol{K}(a), \boldsymbol{I}(a)) \tag{1}$$

where vector $\boldsymbol{I}(a)$ is investment at age a broadly defined. There is no requirement that $\boldsymbol{K}(a+1)$ and $\boldsymbol{K}(a)$ are measured in the same units, or even represent the same skills at different ages. We break $\boldsymbol{I}(a)$ into two components, $\boldsymbol{I_p}(a)$, investment due to the intervention we study, and $\boldsymbol{I_e}(a)$, investment due to other factors that promote learning, including environmental and innate factors. We focus on $\boldsymbol{I_p}(a)$ in most of this paper, but $\boldsymbol{I_e}(a)$ operates in the background.

Knowledge is assumed to be self productive $\boldsymbol{f}_1^{(a)} = \frac{\partial \boldsymbol{K}(a+1)}{\partial \boldsymbol{K}(a)} > 0$ (skill begets skill) and investment is assumed to be productive $\boldsymbol{f}_2^{(a)} = \frac{\partial \boldsymbol{K}(a+1)}{\partial \boldsymbol{I}(a)} > 0$. Static complementarity between skill and investment is found in many studies $\boldsymbol{f}_{1,2}^{(a)} = \frac{\partial^2 \boldsymbol{K}(a+1)}{\partial \boldsymbol{K}(a) \partial \boldsymbol{I}'(a)} > 0$ (those who know more learn more rapidly).[1] Dynamic complementarity captures the

[1] See, e.g., Cunha and Heckman (2007); Cunha et al. (2010); Agostinelli and Wiswall (2021); Cunha et al. (2010); Del Boca et al. (2014) and Cunha and Heckman (2008); Cunha et al. (2010);

notion that investment at age a raises the productivity (rate of learning) of later investment (i.e., $\boldsymbol{I}(a+j)$, $j>0$); that is,

$$\frac{\partial^2 \boldsymbol{K}(a+j+1)}{\partial \boldsymbol{I}(a) \partial \boldsymbol{I}'(a+j)} > 0, \tag{2}$$

so early investment makes later life investments more productive.

Dynamic complementarity arises from three sources: (a) productivity of investment at age a — later starters may get less of a boost in skill by investment at the outset compared to early starters (i.e., there may be "sensitive periods" for investment and/or they may receive less investment); (b) depreciation or appreciation of skills post age a investment; and (c) static complementarity of investment at later age $a+j$ with the stock of skill in the period, which itself is a consequence of early investment.

Dynamic complementarity should be distinguished from the effect on the stock of skills at age a that is a consequence of higher investment in previous periods and higher initial stocks of skills. Dynamic complementarity is a statement about the effect of investment at an age on the *growth* of skills at that age and not about the *level* of skills at the age. We present evidence on both phenomena.

We also present evidence on growth in skills unrelated to any targeted investment strategy: children learn from immersion in environments even without any formal instruction. Learning can arise from imitation, peer effects, parental instruction, and innate growth in skills outside of any experimental intervention. Such growth

Del Boca et al. (2025, 2014); Caucutt and Lochner (2020); Chaparro et al. (2020); Del Bono et al. (2022). Heckman and Mosso (2014) discuss evidence that $\frac{\partial^2 \boldsymbol{f}^{(a)}}{\partial \boldsymbol{K}(a) \partial \boldsymbol{I}'(a)}$ is negative at early ages, but increases with age and becomes positive at later ages.

contributes to initial conditions for dynamic complementarity, and is operative for both treatment and control groups, although possibly in different strengths. It is a possible source of "fadeout" of experimental treatment effects if control groups experience skill growth not directly related to the experiment being examined.

This paper presents nonparametric estimates of these concepts that do not invoke conventional assumptions about the existence of common scales for individual skills (e.g., "human capital" or "ability") that hold across different levels of a skill.[2] Multistage technologies often involve different inputs at different stages. However, it is common in the empirical literature on human development to use achievement test scores as scales of skills assumed to be comparable across people and ages (see, e.g., Todd and Wolpin (2006) who use the number of words known as a measure of human capital. Del Boca et al., 2014 use arbitrarily scaled test scores). In a companion paper, Heckman and Zhou (2026) and in Appendix F, we show that this assumption is questionable.

We develop and apply a strategy to test for the existence of dynamic complementarity using data from a unique home visiting intervention that assesses child skills weekly over a 36-month period of early life (China Development Research Foundation (CDRF), 2018). We exploit exogenous variation in the ages at which children enter the program and a structured schedule of repeated instruction and assessments of age-specific skills that applies to all participants of the same age, irrespective of their previous experience in the program. We examine whether the learning rate of skills is accelerated by higher levels of skills present in each stage of instruction.

[2]See Cunha and Heckman (2008) and Cunha et al. (2010, 2021) for a discussion of this problem. Ben-Porath (1967) is an early example of the use of this widely-used assumption.

We also examine the importance of initial conditions and the quality of home and neighborhood environments in promoting learning.

We find evidence supporting dynamic complementarity, but it does not operate uniformly across skill levels or across ability groups. Normal- and low-ability children display dynamic complementarity, while high-ability children do not, because, in our sample, they generally master the required knowledge at the beginning of each level of instruction. Dynamic complementarity enables low-ability children to partially close the gap with high- and normal-ability children. We also find that children raised by grandparents and in villages with many left-behind children do not learn as rapidly. This suggests that in future applications of the program, those groups should be more intensively targeted. We also find evidence consistent with Heckman and Zhou (2026) that new skills are acquired at different levels of nominally the same skill. This helps explain the "fadeout" claimed to exist in many interventions (see, e.g., Bailey et al., 2020) if different skills are being measured at different ages.

This paper proceeds as follows. In Section I, the concept of dynamic complementarity is defined precisely and decomposed into its constituent parts. We also consider skill growth due to factors other than the intervention studied. Section II discusses our data. In this paper, we largely, but not exclusively, focus our attention on children in the treatment group for whom we have data on weekly skill growth. Section III presents measures of learning and knowledge, and evidence on the growth in skills arising from the intervention we study. Section IV presents evidence on dynamic complementarity. Section V reports evidence on the growth of knowledge as a function of exposure to lessons–a concept distinct from dynamic complementarity. Section VI reports evidence on one component of dynamic complementarity: transmission of skills over the lifecycle, also known as "self productivity". Section VII presents estimates of static complementarity. Section VIII presents evidence on growth in skills unrelated to the interventions studied. Section IX summarizes. Generalizations and additional material are found in our appendix.

I Decomposing Dynamic Complementarity and Analyzing Growth from External Factors Unrelated to the Experiment Being Studied

We decompose dynamic complementarity into three components, which we estimate in our empirical analysis. We also consider growth from learning outside the program. Such learning sets the initial conditions of new entrants into the program and characterizes skill growth in the control group of the experiment studied. In Equation (1), all inputs are assumed to have positive marginal products. Investment promotes the development of skills $\frac{\partial \boldsymbol{f}^{(a)}}{\partial \boldsymbol{I}(a)} \mid_{\boldsymbol{K}(a)=\bar{\boldsymbol{K}}} > 0$. Its productivity depends on the level of the capital stock $\boldsymbol{K}(a)$. Age a^{+} is said to be sensitive to investment relative to a if investment is especially productive at a^{+}:

$$\underbrace{\boldsymbol{f}_2^{(a)}(\boldsymbol{K}(a), \boldsymbol{I}(a)) \mid_{a=a^{+}, \boldsymbol{K}(a)=\bar{\boldsymbol{K}}}}_{\substack{\text{Marginal productivity of} \\ \text{investment at } a^{+}}} > \underbrace{\boldsymbol{f}_2^{(a)}(\boldsymbol{K}(a), \boldsymbol{I}(a)) \mid_{a\neq a^{+}, \boldsymbol{K}(a)=\bar{\boldsymbol{K}}}}_{\substack{\text{Marginal productivity of} \\ \text{investment at } a\neq a^{+}}} . \tag{3}$$ [3]

[3] There may be a range of such values.

Dynamic complementarity arises if investment is productive: $\boldsymbol{f}_2^{(a)}(\boldsymbol{K}(a), \boldsymbol{I}(a)) > 0$ and there is complementarity between the stock of skills and investment $\boldsymbol{f}_{1,2}^{(a)}(\boldsymbol{K}(a), \boldsymbol{I}(a)) > 0$ at age a (greater stocks of skills enhance the productivity of investment at a).

Substituting recursively, we can write the marginal product of investment ($\boldsymbol{I}(a)$) at age a on skill ($\boldsymbol{K}(a+j+1)$) at age $a+j+1$ ($j > 0$) as follows:

$$\frac{\partial \boldsymbol{K}(a+j+1)}{\partial \boldsymbol{I}(a)} = \underbrace{\left[\prod_{m=1}^{j} \boldsymbol{f}_1^{(a+m)}\right]}_{\substack{\boldsymbol{Q}(a+j+1):\\ \text{Transmission of}\\ \boldsymbol{I}(a) \text{ to } \boldsymbol{K}(a+j+1);\\ \text{“Carryover” or}\\ \text{“Self Productivity”}}} \underbrace{\boldsymbol{f}_2^{(a)}(\boldsymbol{K}(a), \boldsymbol{I}(a))}_{\substack{\boldsymbol{P}(a):\\ \textbf{Period } \boldsymbol{a} \textbf{ Marginal}\\ \textbf{Productivity}}}, \tag{4}$$

where $\boldsymbol{P}(a)$ is the marginal productivity of investment at age a and $\boldsymbol{Q}(a+j+1)$ is the transmission of period a investment to skill in period $a+j+1$. It is presumed to be positive. Dynamic complementarity exists if the effect of investment at age a on the productivity of future investments is positive:

$$\begin{aligned} &\frac{\partial^2 \boldsymbol{K}(a+j+1)}{\partial \boldsymbol{I}(a) \partial \boldsymbol{I}'(a+j)} \\ &= \underbrace{\boldsymbol{f}_{1,2}^{(a+j)}(\boldsymbol{K}(a+j), \boldsymbol{I}(a+j))}_{\substack{\textbf{Static Complementarity}\\ \textbf{at age } \boldsymbol{a+j}}} \underbrace{\prod_{\boldsymbol{m=1}}^{\boldsymbol{j-1}} \boldsymbol{f}_{\boldsymbol{1}}^{\boldsymbol{(a+m)}}}_{\boldsymbol{Q}(a+j)} \underbrace{\boldsymbol{f}_{\boldsymbol{2}}^{(a)}(\boldsymbol{K}(a), \boldsymbol{I}(a))}_{\boldsymbol{P}(a)} > 0. \end{aligned} \tag{5}$$

Dynamic complementarity thus depends on: (i) the productivity of investment in period a ($\boldsymbol{P}(a)$), (ii) the appreciation or depreciation of period $\boldsymbol{I}(a)$ investment on skills at $a+j$ ($\boldsymbol{Q}(a+j)$), and (iii) the complementarity between skills and investment in period $a+j$ (i.e., $\boldsymbol{f}_{1,2}^{(a+j)}(\boldsymbol{K}(a+j), \boldsymbol{I}(a+j)))$. $\boldsymbol{Q}(a+j)$ is a diagonal matrix with each row associated with a distinct skill. Initial conditions are $\boldsymbol{K}(a)$.[4]

Investments at a relatively less productive age, say $a' > a$, so $\boldsymbol{P}(a') < \boldsymbol{P}(a)$, everything else equal, have lower levels of associated dynamic complementarity. *Ceteris paribus*, a smaller transmission factor $\boldsymbol{Q}(a+j)$ reduces the magnitude of dynamic complementarity.

Dynamic complementarity (5) is about the *rate of growth* of skills with investment. This is in contrast with the effect of investment on the *level of skills* arising from investment (4).

There are several things to notice about equations (4) and (5): (a) the units in which skills at ℓ and $m (\neq \ell)$, $\boldsymbol{K}(\ell)$ and $\boldsymbol{K}(m)$ are measured, need not be the same, even if they bear the same name in common parlance. They could, in principle, be different types of skills, e.g., spatial cognition for ℓ, verbal cognition for m. (b) In principle, different skills can be acquired at different ages. The skill set may expand with age.[5] The technology of learning $\boldsymbol{f}^{(a)}$ can evolve with age. Equation (1) is consistent with skills of a different nature in each period being transformed into new next period skills.[6] (c) For a model with per period depreciation rate σ for skill k, and no augmentation of skill in the period following investment, $\boldsymbol{Q}_k(a+j) = (1-\sigma)^{a+j}$. $\boldsymbol{Q}_k(a+j)$ measures carry over to $a+j$ of skills acquired at a.

[4] $Q(a+j)$ may be non-diagonal if in transmission stocks cross fertilize. We abstract from this consideration in the text of this paper. We consider cross fertilization in Appendix I.2. Data restrictions limit our ability to estimate cross fertilization in full generality.

[5] Eshaghnia et al. (2025) present evidence on this phenomenon arising from executive functioning skills emerging in adolescence.

[6] This feature has a Hayekian flavor (Hayek, 1941). It may be that capital acquired in periods other than the most recent has independent influences on capital accumulation in a current period, so Equation (1) is not first order Markov (see, e.g., Todd and Wolpin, 2006).

The model of Equation (1) incorporates growth in skills for children in whom no formal investment is also made. Even without formal investment, skills can accumulate through imitation, home environments, peer influences, and genetic influences.[7] These forces also operate on the experimentals and are conjoined with the influence of the intervention studied. We define $\boldsymbol{I_e}(a)$ for environmental variables at age a and examine their influence.

Skills accumulate through investment and environments. Control group children have $\boldsymbol{I_p}(a) = 0$ as do cohorts of children prior to entry in the program. However, other factors may be at work, captured by $\boldsymbol{I_e}(a)$.

The key to the empirical strategy of this paper is that we have samples of children who enter the program at different ages. As a result, prior to entry into the formal program, their skills accumulate via the influence of $\boldsymbol{I_e}(a)$. The skills so accumulated plug into technology (1) at entry into formal treatment. This paper measures their contribution.

II China REACH

The program we analyze, *China REACH*, developed by the China Development Research Foundation (CDRF), extends and applies the widely implemented and prototypical Jamaican Reach Up and Learn home visiting program.[8] Implemented in 2015 by a large-scale randomized control trial, China REACH enrolled over 1,500 subjects (age 6 months-42 months) in 111 villages in Huachi county, Gansu province,

[7]See Cook and Newson (2014) for discussion of the genetic basis of learning following Chomsky.

[8]See, e.g., Grantham-McGregor and Smith, 2016 and Gertler et al., 2014. The original Jamaica team designed China REACH for CDRF.

one of the poorest areas of China.

China REACH is a pair-matched RCT that minimizes the mean square error of estimates (Bai et al., 2021; Bai, 2022). A non-bipartite Mahalanobis matching method was used to pair villages and randomly select one village within a pair into the treatment group and the other village into the control group.[9] Details about the design of the experiment and evidence on the balance of treatment and control groups are reported in Zhou et al. (2024) and Heckman and Zhou (2026). We present additional balance tests in this paper. Our primary analysis in this paper uses only data from the treatment villages, but we also analyze the skill development of control children at a less granular level.

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

Zhou et al. (2024) evaluate the treatment effects of the intervention. Two measurements are collected for both the control and treatment groups at midline and endline of the intervention. This paper analyzes additional data on the weekly growth in skills for the treatment group of the study. Zhou et al. (2024) report that the intervention significantly improves skill development (e.g., language and cognitive, fine motor, and social-emotional skills). To interpret treatment effects, they use item response models to estimate individual latent skills. They decompose treatment effects and find that enhancement of latent skills explains most of the estimated conven-

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

tional treatment effects. We build on their analysis by using their data on treatment and control groups to measure skill growth at a more granular level. Zhou et al. (2023) show that the skill profiles for the growth of skills are similar to those of the original Jamaica Home Visiting program over ages where comparable data exist, suggesting the applicability of our analysis to the original program.

II.1 Enrollment Protocol and Evidence on Balance

All children in selected villages between the ages of 6 and 25 months as of September 2015 were eligible at enrollment. There were only few families with siblings so spillovers to other children are not an issue.[10] Figure 1 gives the enrollment time frame, as well as the timing of measurements for the standard Denver assessment analyzed in Zhou et al. (2024) to estimate program treatment effects. The Denver assessment is a standard inventory of child skills.

All children enter the program at the same time. But children differ in ages at enrollment. We test for balance in the backgrounds of enrolled children within and across villages in the various comparisons we make. We do not reject hypotheses about balance. Different entry-by-age cohorts receive different exposures to the program. However, all caregivers of the children of the same age enrolled in the program treatment group get the *same* lessons across villages. Children are evaluated weekly on their knowledge. $\boldsymbol{I}_p(a)$ (the lesson to the caregiver) is exogenous. Visitors are chosen from the target villages and are essentially homogenous across villages with the same level of education as the mothers visited. They are essentially randomly as-

[10] See Bennhoff et al. (2024) for an analysis of sibling spillovers.

signed.[11] However, the implementation of the lessons received depends on caregivers and home environments. By design, older children at entry do not get the training that earlier entrants receive. The program participation rate is 98% and there is no attrition from the program.

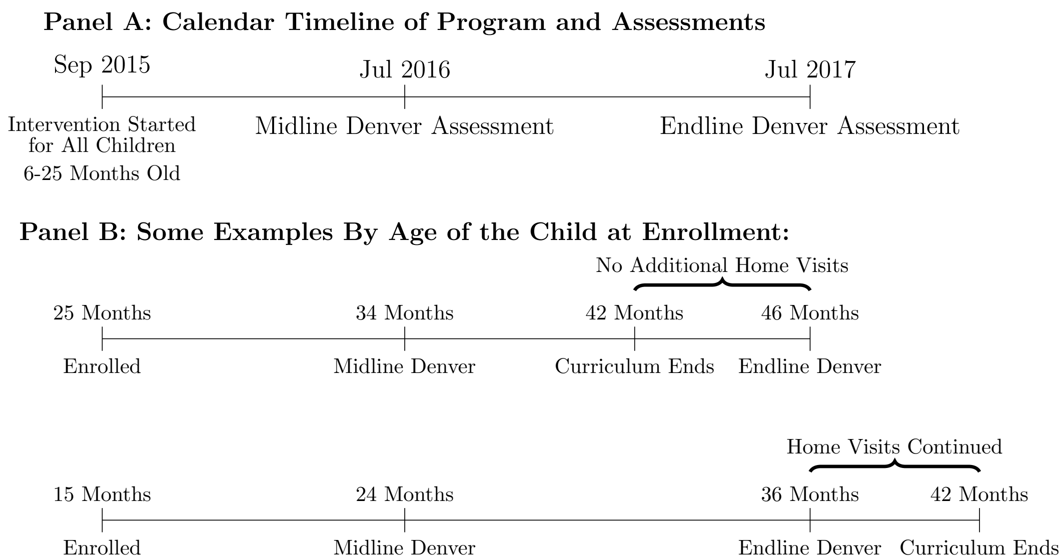
Figure 1 presents the timetable for the program and two examples for different age cohorts of children. Eligible children of different ranges of ages enter the ladder of the program at the rung appropriate to their age. This design creates an essentially random enrollment by age of children of the same backgrounds. Figure 2 shows the balance of the ages of children in each treatment village enrolled in the experiment. The balance is satisfactory.

At midline, a Denver assessment is given to treatment and control children who often differ in age.[12] The treatment group children differ in the lessons received which is determined by their age. This gives a unique variation that makes this study distinctive.

[11] According to the information collected by the CDRF field team, 50 villages out of 55 villages have an average of 9 years of education for home visitors, which is about 90% of all treated villages. For two villages, the average years of education for home visitors are about six years; for two villages, it is about 12 years; and for one village, it is about 14 years. For Pearson χ^2 statistic ($\chi^2(54) = 9.50$), we cannot reject the null hypothesis that all the villages have the same distributions of years of education for the home visitors. When we remove the anomalous villages, we get essentially the same results, as we reported in the main section of this paper. The Pearson statistic is $\chi^2(49) = 12.72$ after removing the anomalous villages. We cannot reject the null hypothesis that caregivers in all the villages have the same years of education. Estimates are essentially the same with and without the inclusion of these villages.

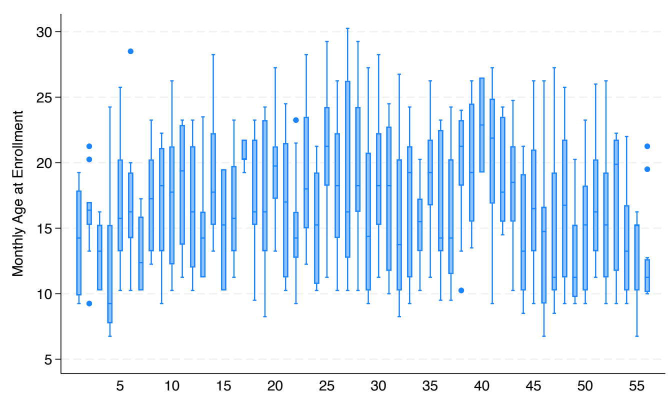
[12] The program is thus *not* adapted to the level of achievement of children in the program. Lessons are identical for all children of a given age. An adaptive design would presumably enhance the performance of children at the cost of tailoring the program curriculum to each child.

Figure 1: China REACH Calendar Time Scales



Notes: 1. This figure presents China REACH calendar time scales and provides two examples. The time points are meant to compare vertically across two panels to show children's monthly age at the point of time during the intervention. 2. Children in the treatment village would complete all home visits when they are 42 months old, and the status of receiving home visits is regardless of the Denver Assessment.

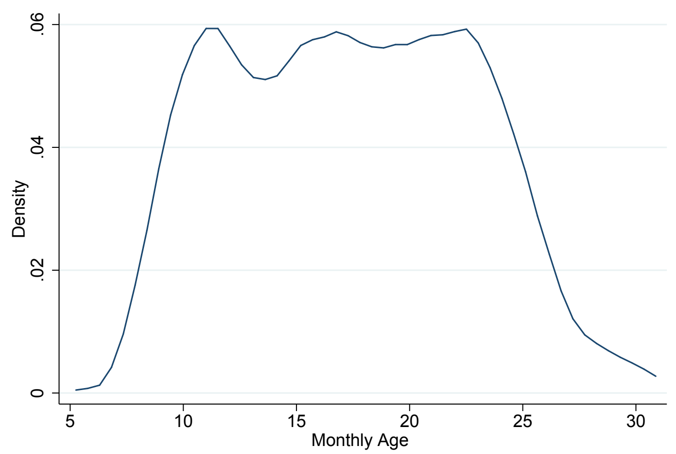
Figure 2: Distribution of Ages at Enrollment Across Treatment Villages



Notes: 1. Box plot showing median, inter-quartile range, and outliers of ages at enrollment across treatment villages. 2. The number of treatment villages' index is presented on the x-axis.

This design is ideal for testing dynamic complementarity, decomposing its components and estimating external learning effects. The entry cohorts are essentially randomly distributed between 10 and 25 months old. Few children older than 25 months are enrolled (see Figure 3).

Figure 3: The Density of Treatment Group by Monthly Age at Enrollment



There is no selection in terms of family investment and endowment across different age cohorts. Parents' years of education and HOME-environment scores are commonly used as measures of family investment. In Table 1, we test whether years of education and baseline HOME-environment scores differ across different enrollment age groups. We find no evidence of differences between years of education and baseline HOME-environment scores across enrollment age groups. Table 1 indicates that, before enrolling in the program, family investment is the same across different age groups within and across villages. We conduct a robustness check by ability groups in Appendix Tables A.1-A.3 and find similar support for the hypothesis of no significant difference in terms of investment before enrollment. Zhou et al. (2024) show balance within pairs by village.

Table 1: Tests of Balance in Parental Education and HOME Scores at Baseline

	Years of Education			HOME Score					
Age Group	Mother	Father	Grandmother	Warmth	Verbal Skills	Hostility	Learning Literacy	Outings	Total*
				Enrollment Age (10-15) vs. (16-20)					
Mean (Age 10-15)	6.889	7.962	2.952	4.803	2.581	4.746	6.411	1.798	23.759
Mean (Age 16-20)	7.319	7.977	3.077	4.835	2.567	4.740	6.441	1.868	23.899
two side p-value	0.150	0.953	0.665	0.630	0.538	0.828	0.828	0.026	0.474
two side step down p-value	0.260	0.982	0.961	0.961	0.928	0.982	0.982	0.018	0.876
p-value (Age 10-15 > Age 16-20)	0.925	0.524	0.668	0.685	0.269	0.414	0.586	0.987	0.763
step down p-value (Age 10-15 > Age 16-20)	0.999	0.974	0.995	0.995	0.774	0.941	0.985	1.000	0.995
N	448	449	439	470	470	470	470	470	470
				Enroll (10-15) vs. (21-25)					
Mean (Age 10-15)	6.889	7.962	2.952	4.803	2.581	4.746	6.411	1.798	23.759
Mean (Age 21-25)	6.743	7.789	2.834	4.772	2.585	4.731	6.443	1.812	23.779
two side p-value	0.627	0.586	0.751	0.693	0.855	0.626	0.815	0.721	0.935
two side step down p-value	0.515	0.452	0.622	0.518	0.796	0.432	0.705	0.605	0.883
p-value (Age 10-15 > Age 21-25)	0.314	0.293	0.376	0.346	0.573	0.313	0.592	0.639	0.533
step down p-value (Age 10-15 > Age 21-25)	0.569	0.601	0.752	0.800	0.903	0.800	0.903	0.903	0.867
N	406	406	400	428	428	428	428	428	428
				Enroll (16-20) vs. (21-25)					
Mean	7.319	7.977	3.077	4.835	2.567	4.740	6.441	1.868	23.899
Mean	6.743	7.789	2.834	4.772	2.585	4.731	6.443	1.812	23.779
two side p-value	0.123	0.547	0.475	0.286	0.423	0.718	0.985	0.145	0.468
two side step down p-value	0.211	0.796	0.796	0.571	0.790	0.817	0.977	0.253	0.796
p-value (Age 16-20 > Age 21-25)	0.061	0.273	0.238	0.143	0.788	0.359	0.507	0.072	0.234
step down p-value (Age 16-20 > Age 21-25)	0.112	0.585	0.561	0.332	0.871	0.668	0.739	0.136	0.561
N	384	385	377	400	400	400	400	400	400

1. Age (10–15) represents children whose monthly ages are between 10 and 15 at enrollment. Age (16–20) represents children whose monthly ages are between 16 and 20 at enrollment. Age (21–25) represents children whose monthly ages are between 21 and 25 at enrollment.

2. Step down p values are constructed by multiple hypotheses between the earlier enrolled group and later enrolled group based on Romano and Wolf (2005a,b).

3. Both step down p-values are conducted by 5000 times of bootstrap and cluster at the village level.

* The sum of each HOME category in this table.

II.2 Program Protocols

The program teaches caregivers to interact with their children through playing games, making toys, singing, reading, and storytelling to stimulate the child's cognitive, language, and socioemotional skill development. The home visit to caregivers is the investment we study ($\boldsymbol{I_p}(a)$). We lack data on the precise way parents act on the information they receive. However, using a rich set of observed caregiver characteristics, we estimate how caregivers with different attributes mediate the impact of home visits on child development. We also study how the home and neighborhood environments affect out-of-program development.

Tasks in four different skill categories (gross motor, fine motor, language, and cognitive) are taught each week. Skills taught are ordered by difficulty levels following profiles developed by Palmer (1971) and Uzgiris and Hunt (1975), henceforth UHP.[13] These scales are widely used in the literature on child development, and are the ones analyzed in this paper. The intervention instructs caregivers at the weekly level on protocols to promote the skills that appear in these scales.[14]

Central to the estimation strategy of this paper is the use of scales of skills that describe levels of knowledge with content that is **the same** within each level and across all children at that level.[15] This is in stark contrast to widely-used test scores, such as the AFQT or IQ tests, with items that are not necessarily comparable across items, except by psychometric fiat.[16] Child skills are assessed weekly. There are

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

[14]Some of these scales also appear in the Denver tests we analyze.

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

[16]See, e.g., Lord and Novick (1968) and Torgerson (1958) for a discussion of the artifice of

monthly assessments of the quality of home visits recorded by supervisors, and data on the quality of home environments are also collected. Home visitors operate in multiple geographically separated villages. We do not analyze gross motor skills in this paper because there is no program treatment effect for those skills (see Zhou et al., 2024).

There are eleven UHP difficulty levels for language skills (see Table 2). The language skills curriculum teaches children to communicate their needs, thoughts, feelings, and ideas in a way that the caregiver can understand. It includes vocalizations, gestures, spoken words, and other signals. Language skill difficulty levels are based on the concepts shown in Table 2. The language skill tasks increase in difficulty with the expectation that the child will learn to identify and use expressive language to indicate understanding. The tasks begin with the baby passively listening as the caregiver makes sounds and speaks. The child then plays a more active role, expected to indicate understanding (receptive language) and use simple gestures to indicate meaning. As understanding and vocabulary increase, the child names more pictures and learns to describe them. Finally, the child learns the names and uses of objects in the child's everyday environment.

Standard psychometric practice imposes a common numerical scale across these levels and claims to quantify the magnitude of each level that can be used to define and construct a cardinal scale of "language comprehension" that can be compared across levels. Careful inspection of the content of each level suggests that they capture very different skills across levels that can be compared only by psychometric

creating conventional test scores.

convention.

Table 2: Overview of Language Task Child Should Perform (Learning Words)

Level 1	Caregiver and baby make sounds to each other to interact
Level 2	Caregiver tells baby the things she does in the house
Level 3	Baby recognizes people's names
Level 4	Baby learns movements that show intimacy: clapping, bye-bye, and thank you
Level 5	Caregiver and child look at the pictures together, and let the child vocalize and touch the pictures
Level 6	Baby recognizes at least one body part
Level 7	The child identifies and/or names ordinary objects
Level 8	The child points to the pictures which are being named, names one or more pictures, mimic the sound of the objects
Level 9	The child points to the pictures which are being named, names two or more pictures, mimic the sound of the objects
Level 10	The child points at 7 or more than 7 pictures and talk about them
Level 11	Teach the child some simple descriptive words and the child names objects at home, and tells the usage of those objects

Table 3: Language Task Content (Learning Words) Level Three

Difficulty Level	Difficulty Level Aim	Month	Week	Learning Materials	Task Aim and Content
Level 3	To teach baby to recognize people's names	9	2	Language – correct nouns:	To teach baby to recognize people's names: Mother teaches baby the names of family members
Level 3	To teach baby to recognize people's names	11	4	Language – correct nouns:	To teach baby to recognize people's names
Level 3	To teach baby to recognize people's names	14	2	Language – correct nouns:	To teach baby to recognize people's names: Mother teaches baby to say the names of the family members
Level 3	To teach baby to recognize people's names	16	3	Language – correct nouns:	To teach baby to recognize people's names: Mother teaches the child to identify the names of family members.
Level 3	To teach baby to recognize people's names	18	1	Language – correct nouns:	To teach baby to recognize people's names: Mother teaches the child the names of family members.

In contrast, Table 3 presents detailed information about the five tasks (and assessments) within difficulty level 3 directed to children 9-18 months old. All tasks relate to the well-defined activity of teaching the baby to recognize people's names and are essentially identical. In addition, all evaluation criteria are the same for each of these tasks. There is no hierarchy of tasks assigned *within* levels. This is a feature of each level of the UHP scale we use. In fact, some tasks in each level are exactly repeated while others are only slightly altered. Appendix C documents the task content for each difficulty level for all of the skills taught and examined. The home visitors give lessons and perform assessments of the lessons on the children. The fact that the skills taught and assessed *within levels* are essentially identical gives us a natural scale comparable within levels and comparable across different persons, but not necessarily across levels.

II.3 Timing of Instruction

Language instruction begins at the outset. Only when children reach level 6 in language does instruction in cognition and fine arts begin. This compromises our ability to estimate cross effects of skills prior to that age and limits our ability to estimate the general technology (1) at all ages. When multiple skills are being taught, children are in different instructional stages within levels, compromising measures of complementary skills. This affects our strategy of estimating the components of (5), which we attempt in later sections of the paper.

Table 4 presents the monthly age range children would receive each skill level. One thing needs to be noticed: the difficulty levels of each skill are not comparable in terms of content or the monthly age at which children are exposed to them. Appendix C documents the monthly ages and the task content of each level and skill. Children who enrolled at later monthly ages did not miss as many Cognitive/Fine Motor tasks compared to Language tasks. Children enrolled at Language level 6/Cognitive level 5/Fine Motor level 2 miss 18 Cognitive tasks, 8 Fine Motor tasks, compared to 44 Language tasks respectively. The timing and intensity of early investment affect our ability to estimate Equation (1) in full generality.

Table 4: Monthly Age Ranges in the Curriculum Design

	Language			Cognitive			Fine Motor		
Difficulty Level	First Task	Last Task	Number of Tasks	First Task	Last Task	Number of Tasks	First Task	Last Task	Number of Tasks
1	6.00	8.50	6	10.25	15.25	7	12.50	20.25	6
2	6.75	20.00	20	16.00	22.00	7	21.00	22.50	2
3	9.50	18.25	5	20.50	21.25	2	23.50	29.75	6
4	10.00	18.50	7	21.75	22.25	2	30.25	34.25	6
5	10.50	15.50	6	23.00	32.75	11	36.00	37.50	3
6	10.75	25.25	10	26.00	36.00	9	38.50	41.75	4
7	19.25	31.50	6	23.75	33.50	10	40.25	42.75	3
8	21.75	40.75	10	26.25	35.75	7			
9	26.00	42.75	7						
10	26.00	39.00	5						
11	34.00	42.50	9						

Notes: This table shows the monthly age at which children would receive each skill by curriculum design. The difficulty levels of each skill are not comparable in terms of content or the monthly age children exposed.

III Measuring Learning

To understand the structure of the data analyzed, define $\mathcal{S}$ as the set of skills taught. Let $\ell(s, a)$ be the level of skill s taught at age a. Within each level, the skill taught and assessed is identical, as documented for language learning in Table 3. Mastery of skill s at level ℓ at age a is characterized by a latent variable crossing a threshold:

$$D(s, \ell, a) = \begin{cases} 1 & K(s, \ell, a) \geq \bar{K}(s, \ell) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $D(s, \ell, a)$ records mastery (or not) of skill s at a given level at age a. $K(s, \ell, a)$ is a latent variable reflecting knowledge of skill s at the level ℓ at age a. $\bar{K}(s, \ell)$ is the minimum latent skill required to master the task at difficulty level ℓ.[17] Define

[17] This characterization is consistent with the classical IRT model (Lord and Novick, 1968) and models of discrete choice (Thurstone, 1927; McFadden, 1981).

$\underline{a}(s,\ell)$ as the first age at which skill s is taught at level ℓ in the intervention studied, and let $\bar{a}(s,\ell)$ be the last age at which it is taught at level ℓ. For level ℓ of skill s, indicators of knowledge in a spell are elements of $\left\{D(s,\ell,a)\right\}_{\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}$.

III.1 Measures of Learning and Knowledge

In our samples, knowledge of skill s (passing rate) at level ℓ at age a is the mean across individuals of $D(s,\ell,a)$, the mean passing rate for the item. $Pr(D(s,\ell,\bar{a}(s,\ell))=1)$ is a measure of final skill s level attainment in level ℓ. ***Time to first mastery*** is $d(s,\ell)=\hat{a}(s,\ell)-\underline{a}(s,\ell)$, where for each s and ℓ, $\hat{a}(s,\ell)=\min_a\{D(s,\ell,a)=1\}_{a=\underline{a}(s,\ell)}^{\bar{a}(s,\ell)}$. This is a measure of learning speed. The number of attempts to achieve first mastery across levels of nominally the same skill is a measure of knowledge of that skill (van der Linden, 2016). Arguably, it is a measure of ability.

III.2 Patterns of Learning

Figure 4 characterizes the growth of knowledge in language, cognitive, and fine motor skills.[18] Average (across people) passing rates by age within each difficulty level for language and cognitive tasks increase with the number of lessons, a pattern consistent with learning. When individuals transition to higher difficulty levels, initial age-specific passing rates decline at entry into the level. This is consistent with the notion that new skills are taught at each level.[19] After initial declines, age-specific

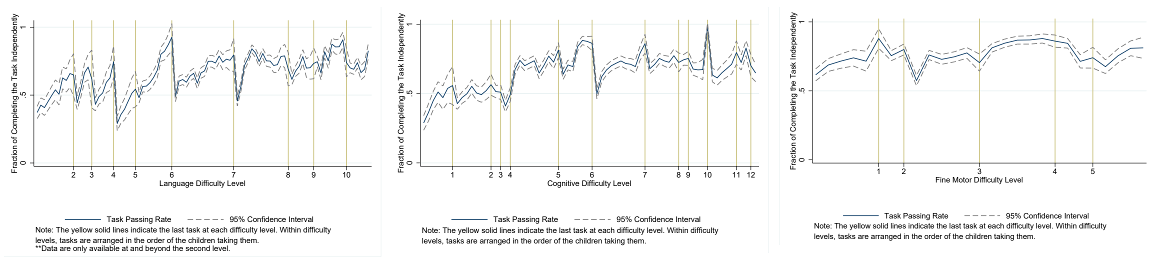
[18]We also measure gross motor skills, but they are not affected by the intervention (Heckman and Zhou, 2026), so we do not systematically analyze them here.

[19]Alternatively, this might arise if the thresholds of assessments for the same nominal skill increase at all transitions across levels, an interpretation we view as implausible.

passing rates within levels increase as learning of the new skill ensues. At most levels of fine motor skills, there is—at best—modest learning. Access to detailed weekly data enables us to determine at what stages learning occurs, and at what rate.

Figure 4: Average Task Passing Rates by Order and Level

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

It is informative to examine the rate of learning across ability levels. We use a measure of learning success on the *initial* level attempted. Appendix D uses a measure of ability defined as an average of success on first attempts across levels and reaches results similar to the results we report here.[20]

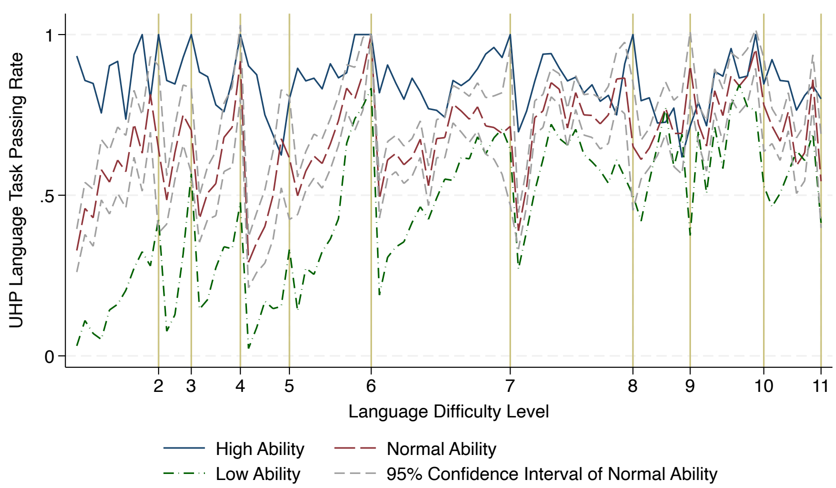
Specifically, in the text, we analyze growth in skills by initial learning speed. Our measure of low ability is that the average passing rate of tasks in the first *three* months after enrollment is less than 20%. Normal and high ability learning are defined as the average passing rate of first *three* months' tasks is higher than 20% and less than 80%, and high ability is higher than 80%, respectively. We use a measure based on the first three months to reduce any measurement errors while still

[20]This is not surprising in light of the high correlation in learning rates across levels.

capturing their baseline performance. The correlation between the average measure across levels presented in Appendix D and our ability measure of low ability in the text is 0.46.

Figure 5 plots UHP language learning curves across levels by ability. First, high ability children have the highest performance across levels, normal the second, and low the third. Second, high ability children have a similar performance as normal ability children in the later skill levels, especially after level 8. Appendix E documents the learning curves for cognitive and fine motor skills in both the initial ability measure and the ability measured across all levels.

Figure 5: **Average Language Task Passing Rates by Ability Level**



Notes: 1. The yellow solid lines indicate the last task at each difficulty level. Within difficulty levels, tasks are arranged by the order of the children taking them.

There are two important takeaways. The first takeaway is that, in later levels, there is strong evidence that lower performers at earlier stages can catch up in learning. The distinction between ability groups is sharp at earlier levels but gradually narrows as children progress across levels. A lot of children who were classified as normal ability even become high ability children.

A second takeaway is that low ability is a relatively stable category in early childhood learning. The patterns discussed in Appendix D show that, with other measures of low ability, the group is a relatively stable category.

In Appendix F, we test for comparability of skills across levels using a standard IRT framework. Consistent with the analysis of Heckman and Zhou (2026), we reject the hypothesis of a common language skill scale across levels, same for cognitive skill. We fail to reject the hypothesis for fine motor skills.

IV Testing Dynamic Complementarity

The hypothesis of dynamic complementarity is that at age $a + j + 1$ for investment at age a

$$\boldsymbol{H}_0 : \frac{\partial^2 \boldsymbol{K}(a + j + 1)}{\partial \boldsymbol{I}(a) \partial \boldsymbol{I}'(a + j)} > 0 \tag{7}$$

for $j > 0$.

The test is meaningful only if we control for initial conditions and factors that promote out-of-program learning. Children enter with different levels of exposure to the intervention curriculum and with different levels of pre-intervention learning. Any meaningful test of $\boldsymbol{H}_0$ requires that we standardize starting points. In this section, we present direct evidence on $\boldsymbol{H}_0$. We report estimates of the components of (5) in subsequent sections. Instead of testing the hypothesis over an aggregate of all skills and levels, we test separately by skill and level to examine the process of learning in detail.

Skills evolve across levels and are not strictly comparable. We have data on the knowledge of items for each skill at each level for children who participate in the treatment group of the intervention. We test hypothesis $\boldsymbol{H_0}$ in this section. In later sections, we attempt to estimate the components of Equation (5) that define dynamic complementarity, but face difficulties with data availability discussed in Section II.3. We first outline our empirical strategy and then apply it to the data.

Define $\boldsymbol{Z}$ as background variables, including parental and caregiver characteristics and other attributes of the learning environment. In principle, we can measure knowledge of skill s at level ℓ at each age by constructing sample counterparts to $\Pr(D(s,\ell,a) = 1|\boldsymbol{Z})$.

We conduct an age-by-age (task-by-task) analysis within levels. Heckman and Zhou (2026) and Appendix F show that measurement error is generally not a serious issue with these data across all skills and skill levels.

A direct test of $\boldsymbol{H_0}$, level-by-level and for each skill, is to tabulate the growth of skills at age a for each level as a function of the history of investment up to a. Our measure of investment at age $a' < a$ is exposure to the program up to a'. We test $\boldsymbol{H_0}$ by comparing the rate of growth of skill as a function of exposure (lessons received). In this section, we conduct a nonparametric version of this test conditioning on family, village, and other background variables that may shape initial conditions and out-of-program learning.[21] This is our main test of dynamic complementarity.

Another approach to testing this hypothesis is to assume that skills accumulate (and are commensurate over levels), so that, in principle, we can update the lessons received and knowledge acquired at each age within levels. In this application, this approach produces very noisy estimates. Instead, for each skill s and level ℓ, we use knowledge inherited from level $\ell-1$ at the start of level ℓ as a measure of the stock of knowledge accumulated through level $\ell-1$.[22] Our measure is $\Pr(D(s,\ell-1,\bar{a}(\ell-1)) = 1|\boldsymbol{Z})$.

It captures knowledge at the end of the skill s level $\ell-1$.[23] More precisely, for a separable in K technology, this approach to testing $\boldsymbol{H_0}$ is as follows. We can proximate the technology for skill s, at age $a+1$ with level ℓ by

$$K(s,\ell,a+1) \doteq F^{(s,\ell)}(K(s,\ell-1,\bar{a}(s,\ell-1)), \boldsymbol{I}(s,\ell,a), \boldsymbol{Z})$$

for each $a(s,\ell)$ in level ℓ. This captures the growth by age of skill s at level ℓ, assuming that the impact of prior investments is captured by lagged $K(\cdot)$. This assumption is consistent with our Markovian assumption about the technology (1) if we impose separability in K.

[21] Campbell and Ramey (1994) test for dynamic complementarity using experimental variation. They randomize entry to treatment in a first stage, then re-randomize in the middle of the experiment to test if those receiving two doses of treatment do better than those who receive only one dose at the end of the experiment and report evidence supporting it. Later work by Meghir et al. (2023) applies this idea to another program, but finds no evidence of dynamic complementarity. Our approach tests dynamic complementarity at a much more granular (weekly) level.

[22] This is consistent with Hayek's capital theory that capital of one form in one period can be transformed into capital in a later period, although the two capitals can be very different (Hayek, 1941).

[23] In practice, we use an average over the last few ages in the previous interval.

At the junction between levels ℓ and $\ell+1$, for all s, $K(s, \ell+1, a(s, \ell+1)) = F^{(s,\ell)}(K(s, \ell, \bar{a}(s, \ell)), \boldsymbol{I}(s, \ell, a), \boldsymbol{Z})$. Investment is measured by lessons in the curriculum. This second test captures static complementarity and also dynamic complementarity under the Markov assumption. We report tests-based on it below, but under the restrictive assumptions just stated. Our main test does not require the Markov assumption.

IV.1 Levels of Exposure to Learning

Our principal test investigates the effect of investment at all previous levels (i.e., earlier exposures) on the learning speed at the current level (i.e., productivity of investment). The design of the China REACH curriculum allows us to investigate this issue because our measure of performance of the same task is fully comparable across all members of the treatment group. They receive the *same* lesson at the same monthly age. We measure children's performance after each lesson. Furthermore, treatment group children enroll at different ages. Table 5 shows the percentage of treatment children who skip levels at enrollment due to their age at enrollment. In our treatment sample, almost all children receive lessons starting from level 6, but the heterogeneity in exposures prior to level 6 gives considerable variability in earlier investment. Our multistage technology does not require us to assume a common scale of skills across levels of the same skill. Table 5 gives the distribution of levels skipped at the time of enrollment in our sample. 43% take the entire curriculum. Many enter at later stages of the curriculum ladder. Virtually all subjects receive lessons at levels 6 and above.

Table 5: Distribution of Skipped UHP Levels at Enrollment (Language)

Skipped Levels	Frequency	Percent
0	304	43.24%
1	131	18.63%
2	3	0.43%
3	47	6.69%
4	79	11.24%
5	108	15.36%
6	30	4.27%
7	1	0.14%

Notes: Treatment group children would skip UHP levels that are designed for younger ages than their age at enrollment. Home visitors would not recover the skipped UHP levels.

IV.2 Evidence on Dynamic Complementarity

In this section, we separate treatment group children by their enrollment status: children who do not skip any levels of enrollment and children who enroll at level 6 and have no prior investment. Doing so conditions on children's exposure in the earlier levels. In Table 6, we conduct balance tests on our background $\boldsymbol{Z}$ variables between two entry cohorts with different program exposure. Table 7 presents the balance test of baseline characteristics between children with full exposure and children enrolled at Level 6. These tests show that balance is found in both comparisons.

Table 6: Balance of $\boldsymbol{Z}$ by Program Exposure

Cohort by Program Exposure	Grandparents	% Left-Behind in the Village	Male
Mean (Full Exposure)	0.087	0.094	0.524
Mean (Enrolled at Language Level 6)	0.081	0.093	0.547
two side p-value	0.585	0.937	0.527
two side step down p-value	0.746	0.911	0.746
N	711	711	711

Notes: 1. Full Exposure stands for children who enrolled before difficulty level 6. Children enrolled at level 6 stand for children who skipped all UHP levels prior to level 6. 2. Step down p values are constructed by multiple hypotheses between the earlier enrolled group and later enrolled group based on Romano and Wolf (2005a,b). 3. Both step-down p-values are calculated using 5,000 bootstrap iterations and clustered at the village level.

Table 7: Balance of Parental Education and HOME Scores at Baseline by Program Exposure

Cohort by Program Exposure	Years of Education			HOME Score					
	Mother	Father	Grandmother	Warmth	Verbal Skills	Hostility	Learning Literacy	Outings	Total*
Mean (Full Exposure)	6.854	7.761	3.046	4.840	2.582	4.730	6.433	1.818	23.821
Mean (Enrolled at Language Level 6)	7.161	8.030	2.886	4.764	2.580	4.752	6.397	1.820	23.749
two side p-value	0.141	0.272	0.565	0.186	0.907	0.431	0.718	0.949	0.680
two side step down p-value	0.208	0.444	0.864	0.264	0.977	0.707	0.922	0.977	0.908
N	664	666	651	700	700	700	700	700	700

Notes: 1. Full Exposure stands for children who enrolled before difficulty level 6. Children enrolled at level 6 stand for children who skipped all UHP levels prior than level 6. 2. Step-down p values are constructed by multiple hypotheses between the earlier enrolled group and later enrolled group based on Romano and Wolf (2005a,b). 3. Both step-down p-values are calculated using 5,000 bootstrap iterations and clustered at the village level.
The sum of each HOME category in this table.

Table 8: Time to First Mastery by Enrollment Cohort (Language)

	UHP Language Levels ℓ				
$d(s,\ell)$	$\ell=7$	$\ell=8$	$\ell=9$	$\ell=10$	$\ell=11$
Full Exposure of Curriculum	2.487	2.079	1.496	1.404	1.348
Enrolled at Language Level 6	2.518	2.367	1.686	1.454	1.698
p-value	0.870	**0.024**	**0.017**	0.608	**0.000**
stepdown p-value	0.864	**0.059**	**0.059**	0.827	**0.000**

Notes: 1. This table illustrates the learning speed within the same UHP Language level, conditional on enrollment cohorts. Time to first mastery is the number of trials a child takes until the first success at each difficulty level during the intervention for each skill type. 2. Children with full exposure to the curriculum stand for children who did not skip any UHP Language levels prior to level 6. Children enrolled at Language level 6 stand for children who skipped all UHP Language levels prior to level 6. 3. Stepdown p-values are constructed by multiple hypotheses between children with full exposure versus children enrolled at Language level 6 across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

Table 8 summarizes our main analysis. We compare children's *learning speed* using the number of trials they take to achieve their first success at the same UHP level. At UHP Language levels 8, 9, and 11, children with full exposure to the curriculum require significantly fewer trials until their first success within each level. Otherwise similar children with more early investment are learning more than those without early investment at later levels.

Tables 9 and 10 present evidence on dynamic complementarity for cognitive and fine motor skills. An important feature of our data is that the difficulty levels of each skill are not comparable in terms of content or the monthly age children exposed.[24] The evidence on dynamic complementarity for cognitive skills is statistically significant for most levels. Fine motor skill does not show a similar pattern. As Table 4 indicates, children who enrolled at later monthly ages do not miss as many cognitive/fine motor tasks compared to language tasks. The children enrolled at Language level 6/Cognitive level 5/Fine Motor level 2 miss 18 cognitive tasks, 8 fine motor tasks, compared to 44 language tasks respectively. The magnitude of early investment affects estimates of dynamic complementarity.

Table 9: Time to First Mastery by Enrollment Cohort (Cognitive)

	UHP Cognitive Levels ℓ			
$d(s,\ell)$	$\ell=5$	$\ell=6$	$\ell=7$	$\ell=8$
Full Exposure of Curriculum	1.748	1.679	2.148	1.579
Enrolled at Cognitive Level 4	2.259	2.065	2.321	1.867
p-value	**0.003**	**0.001**	0.285	**0.007**
stepdown p-value	**0.028**	**0.008**	0.434	**0.043**

Notes: 1. This table illustrates the learning speed within the same UHP Cognitive level, conditional on enrollment cohorts. Time to first mastery is the number of trials a child takes until the first success at each difficulty level during the intervention for each skill type. 2. Children with full exposure to the curriculum stand for children who did not skip any UHP Cognitive levels prior to level 4. Children enrolled at Cognitive level 4 stand for children who skipped all UHP Cognitive levels prior to level 4. The levels at which children all receive the same training vary by level. 3. Stepdown p-values are constructed by multiple hypotheses between children with full exposure versus children enrolled at Cognitive level 4 across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

[24] Appendix C documents the monthly ages and the task content of each level and skill.

Table 10: Time to First Mastery by Enrollment Cohort (Fine Motor)

	UHP Fine Motor Levels ℓ				
$d(s,\ell)$	$\ell=3$	$\ell=4$	$\ell=5$	$\ell=6$	$\ell=7$
Full Exposure of Curriculum	1.674	1.308	1.189	1.571	1.357
Enrolled at Fine Motor Level 2	1.810	1.403	1.171	1.505	1.365
p-value	0.136	0.225	0.741	0.556	0.956
stepdown p-value	0.445	0.567	0.926	0.896	0.948

Notes: 1. This table illustrates the learning speed within the same UHP Fine Motor level, conditional on enrollment cohorts. Time to first mastery is the number of trials a child takes until the first success at each difficulty level during the intervention for each skill type. 2. Children with full exposure to the curriculum stand for children who did not skip any UHP Fine Motor levels prior to level 2. Children enrolled at Fine Motor level 2 stand for children who skipped all UHP Fine Motor levels prior to level 2. 3. Stepdown p-values are constructed by multiple hypotheses between children with full exposure versus children enrolled at Fine Motor level 2 across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

Table 11: Effects of $\boldsymbol{Z}$ on Time to First Mastery (Language)

	UHP Language Levels ℓ				
$d(s,\ell)$	$\ell=7$	$\ell=8$	$\ell=9$	$\ell=10$	$\ell=11$
Grandparent	0.417	0.532*	0.273	0.378**	0.454***
	(0.331)	(0.300)	(0.171)	(0.159)	(0.164)
% Left-Behind in the Village	5.819*	3.668	0.459	1.941	1.500
	(2.922)	(2.446)	(0.867)	(1.189)	(1.190)
Father's Years of Education	0.066	0.049	0.009	0.008	0.018
	(0.045)	(0.041)	(0.006)	(0.006)	(0.016)
Mother's Years of Education	-0.061	-0.051	-0.011*	-0.005	-0.018
	(0.045)	(0.041)	(0.006)	(0.006)	(0.017)
Male	-0.139	0.092	0.013	-0.042	-0.052
	(0.186)	(0.127)	(0.079)	(0.113)	(0.070)
Grandparent stepdown p-value	**0.072**	**0.024**	**0.024**	**0.008**	**0.008**

Notes: 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report average marginal effects from our fitted OLS. Standard errors clustered at the village level are reported in parentheses.
2. The vector of Z includes gender, type of primary caregiver, parents' years of education, and fraction of left-behind children at the village level. 3. Stepdown p-values are constructed by multiple hypotheses between children's primary caregivers, who are grandparents or parents, across UHP levels based on Romano and Wolf (2005a,b) by 250 times of bootstrap.

Table 11 presents coefficients on $\boldsymbol{Z}$ of an OLS regression on children's time to first mastery across UHP Language levels. Time to first mastery is measured by the number of trials children take until reaching their first correct answer.

Grandparents are less educated and, as caretakers, are negatively associated with children's learning speed. Children raised by grandparents require significantly more tasks to get the first correct answer. We report similar results for other skills in Appendix G.

Table 12: **Initial Level of Ability:** Time to First Mastery Among Low Ability Group by Enrollment Cohort (Language)

	UHP Language Levels ℓ				
$d(s,\ell)$	$\ell=7$	$\ell=8$	$\ell=9$	$\ell=10$	$\ell=11$
Full Exposure of Curriculum	4.336	2.486	1.655	1.229	1.640
Enrolled at Language Level 6	4.753	4.073	2.107	1.887	2.579
p-value	0.413	**0.000**	**0.001**	**0.009**	**0.000**
stepdown p-value	0.417	**0.000**	**0.007**	**0.016**	**0.000**

Notes: 1. This table illustrates the learning speed within the same UHP level, conditional on enrollment cohorts. Time to first mastery is the number of trials a child takes until the first success at each difficulty level during the intervention for each skill type. 2. Children with full exposure to the curriculum stand for children who did not skip any UHP levels prior to level 6. Children enrolled at level 6 stand for children who skipped all UHP levels prior to level 6. 3. Stepdown p-values are constructed by multiple hypotheses between children with full exposure versus children enrolled at level 6 across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

Tables 12-14 provide further evidence on learning focusing only on the low ability group. Among the low-ability group in both enrollment cohorts, the difference in learning rates at later levels is more pronounced. Stratifying by ability emphasizes the importance of investing early in low-ability children.[25] Disadvantaged children benefit the most in early stages through dynamic complementarity and reduce the

[25]Low-ability children are more often raised by low-educated grandparents and living in villages with other left-behind children.

gap at later stages with higher ability children and children with better social learning environments.

Table 13: **Initial Level of Ability:** Time to First Mastery Among Low Ability Group by Enrollment Cohort (Cognitive)

	UHP Cognitive Levels ℓ			
$d(s, \ell)$	$\ell = 5$	$\ell = 6$	$\ell = 7$	$\ell = 8$
Full Exposure of Curriculum	2.423	2.060	2.812	2.047
Enrolled at Cognitive Level 4	4.516	2.735	3.559	2.788
p-value	**0.000**	**0.018**	0.103	**0.014**
stepdown p-value	**0.002**	0.103	0.317	**0.091**

Notes: 1. This table illustrates the learning speed within the same UHP level, conditional on enrollment cohorts. Time to first mastery is the number of trials a child takes until the first success at each difficulty level during the intervention for each skill type. 2. Children with full exposure to the curriculum stand for children who did not skip any UHP levels prior to level 5. Children enrolled at level 5 stand for children who skipped all UHP levels prior to level 5. 3. Stepdown p-values are constructed by multiple hypotheses between children with full exposure versus children enrolled at level 5 across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

Table 14: **Initial Level of Ability:** Time to First Mastery Among Low Ability Group by Enrollment Cohort (Fine Motor)

	UHP Fine Motor Levels ℓ				
$d(s, \ell)$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	$\ell = 7$
Full Exposure of Curriculum	1.959	1.446	1.321	1.643	1.667
Enrolled at Fine Motor Level 2	3.293	1.897	1.478	1.765	1.692
p-value	**0.000**	**0.069**	0.385	0.689	0.968
stepdown p-value	**0.002**	0.290	0.776	0.922	0.976

Notes: 1. This table illustrates the learning speed within the same UHP level, conditional on enrollment cohorts. Time to first mastery is the number of trials a child takes until the first success at each difficulty level during the intervention for each skill type. 2. Children with full exposure to the curriculum stand for children who did not skip any UHP levels prior to level 3. Children enrolled at level 3 stand for children who skipped all UHP levels prior to level 3. 3. Stepdown p-values are constructed by multiple hypotheses between children with full exposure versus children enrolled at level 3 across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

Appendix H conducts the same analysis using a global measure of low ability computed as an average across levels. We reach the same conclusion using either measure. Dynamic complementarity is most pronounced among low-ability children. Children who do poorly at earlier levels have much higher productivity at the later levels when given earlier investment. It is crucial to invest in low-ability children early to let them catch up with other children.

We present additional evidence on dynamic complementarity assuming a Markovian structure in Section VII. It is also a test of static complementarity without invoking the Markovian structure.

V Impacts of Exposure to the Program on Levels of Knowledge

Dynamic complementarity should be distinguished from the growth of the stock of skills due to previous exposure to the intervention. This distinction is captured by the contrast between expression (5) and expression (4). Both expressions contain common ingredients, but dynamic complementarity has the additional component of static complementarity. We present evidence on each component of these expressions. This section presents evidence on expression (4), isolating its components and those of (5) in Section VII.

Let $a_j^*(s,\ell)$ be the age of entry into the program for agent j at level ℓ for skill s. $a_k^*(s,\ell)$ is the age of entry into the program for agent k. We measure knowledge of child j at $a(s,\ell)$ (represented as $K_j(a,s,\ell)$). For $a \neq a'$, we use the first task passing rates as measures of knowledge for the two agents. We compute $E(K_j(a,s,\ell) \mid a(s,\ell) > a_j^*(s,\ell))$ and $E(K_k(a,s,\ell) \mid a(s,\ell) > a_k^*(s,\ell))$. For $a_k^*(s,\ell) > a_j^*(s,\ell)$, Equation (4) implies the hypothesis that

$$\boldsymbol{H_1}: E(K_k(a,s,\ell) \mid a(s,\ell) \geq a_k^*(s,\ell)) < E(K_j(a,s,\ell) \mid a(s,\ell) \geq a_j^*(s,\ell)),$$
$$\text{all } (s,\ell) \text{ in the set intersection } \{(s,\ell) | a(s,\ell) \geq a_k^*(s,\ell) \text{ and } a(s,\ell) \geq a_j^*(s,\ell)\}.$$

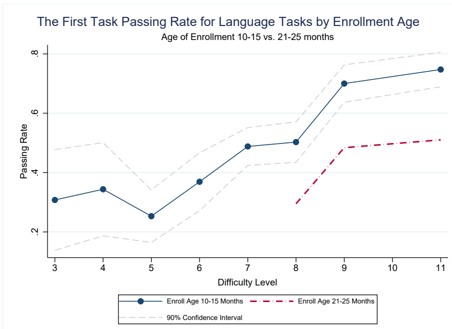
To test hypothesis $\boldsymbol{H_1}$, skill-by-skill and level-by-level within skill, we first categorize three groups based on children's monthly ages at enrollment (i.e., age 10–15, age 16–20, and age 21–25). We then compare children's passing rates on the first tasks across the three groups.

We conduct tests at higher difficulty levels (level 7 and above) to avoid a potential problem arising from age heterogeneity in our data. In the early stages of the experiment, children enter difficulty levels with different levels of experience in the program. Because we group children by the age of enrollment (i.e., 10–15 months, 16–20 months, and 21–25 months), the measures for the first six difficulty levels can be affected by age differences. For example, if child A enrolls in the program at ten months of age, all measures of the child's task performance are initially evaluated at difficulty levels consisting of tasks designed for ten-month-old children, but if child B enrolls in the program at 18 months, the measures are evaluated at difficulty levels consisting of tasks designed for 18-month-olds. Table 4 documents the sample monthly ages for the first and last tasks at each level.

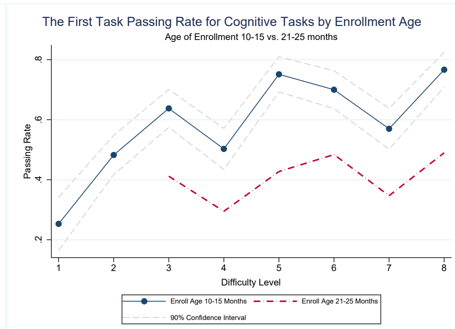
Tests of $\boldsymbol{H_1}$ should be conducted using measures evaluated when children are the same age but have different lengths of program exposure. Thus, we standardize by age using measures for higher difficulty levels (e.g., from difficulty level 7), which are

Figure 6: The Passing Rate on the First Task for Language and Cognitive Tasks by Level and Enrollment Age

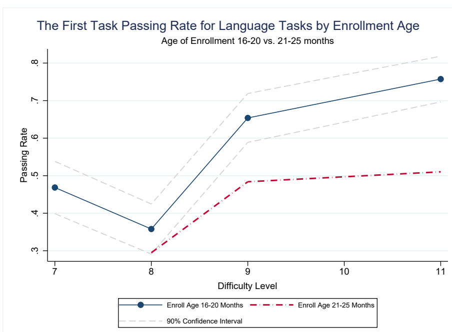
(a) Language: Age (10–15) vs. Age (21–25)



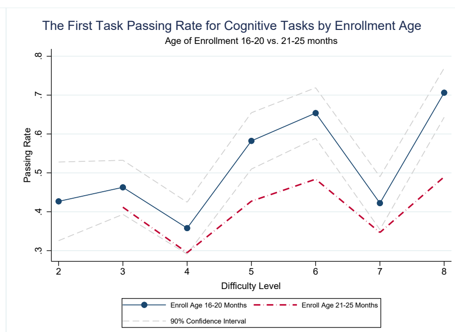
(b) Cognitive: Age (10–15) vs. Age (21–25)



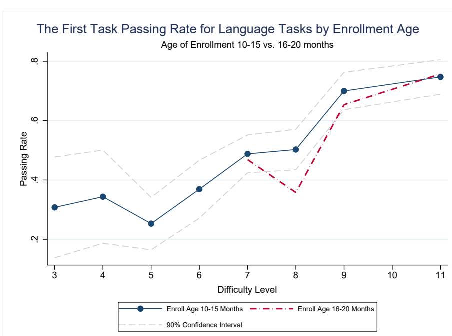
(c) Language: Age (16–20) vs. Age (21–25)



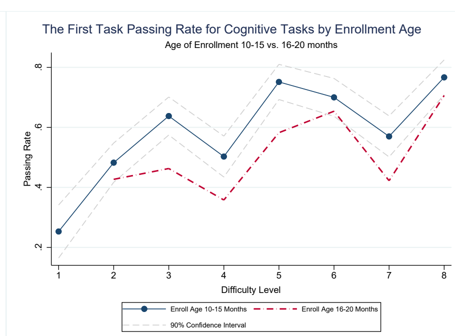
(d) Cognitive: Age (10–15) vs. Age (21–25)



(e) Language: Age (10–15) vs. Age (16–20)



(f) Cognitive: Age (10–15) vs. Age (21–25)



Notes: The blue solid line and red dashed line indicate the first task passing rate at the given difficulty level for children who enrolled in different age groups, respectively.

more suitable for us to accurately examine growth of skill effects because nearly all participants have enrolled at level 6 and above.

Figure 6 plots the passing rates on the first task for language and cognitive skills for children enrolled early and late. Early starters who receive more instruction have a higher level of knowledge than late starters, as predicted by Equation (4).

VI Estimating and Testing Dynamic Transmission

In this section, we attempt to decompose dynamic complementarity into its constituent parts. We first estimate the transmission term (i.e., $\boldsymbol{Q}(a+j)$) in Equation (5), assumed diagonal, consistent with a separable in $\boldsymbol{K}$ technology.[26] Skills evolve across levels. We have data on knowledge of items for each skill at each level for children who participate in the intervention. Our goal is to estimate the components of Equation (5) that define the transmission term (i.e., $\boldsymbol{Q}(a+j)$).

The transmission term in Equation (5), $\boldsymbol{Q}(a+j)$, measures the transmission of accumulated capital across levels for different skills. Within a level, all children receive the same exposure to knowledge, although its impact may depend on the quality of the caregiver, which we measure and place in $\boldsymbol{Z}$. We approximate the term of $\boldsymbol{Q}(a+j)$, from level $\ell-1$ to level ℓ by the following:

$$F_1^{s,\ell} = \frac{\partial K(s,\ell,\bar{a}(\ell,s),\boldsymbol{Z})}{\partial K(s,\ell-1,\bar{a}(\ell-1,s),\boldsymbol{Z})}. \tag{8}$$

The growth in capital from level to level. To estimate transmission from level ℓ to level $\ell+k$, we calculate the joint product term $\prod_{j=\ell}^{\ell+k} F_1^{s,j}$. For each skill s and

[26] We invoke separability in light of our discussion of data limitations in Section II.3.

level ℓ, we use knowledge inherited from $\ell - 1$ at the start of level ℓ as a measure of the stock of knowledge.[27] Our measure for $K(s, \ell, \bar{a}(\ell, s))$ is $\Pr(D(s, \ell, \bar{a}(\ell)) = 1|\boldsymbol{Z})$. This measures knowledge at the end of ℓ for skill s.

VI.1 Estimating $Q(s, a + j, a)$

In this section, we present empirical estimates of the transmission term $F_1^{s,\ell}$ at each level for each skill s. We use the child's last task performance of the current skill level ℓ and child's last task performance of the previous skill level $(\ell - 1)$ and condition on $\boldsymbol{Z}$ in a logit regression[28] to compute the derivative of Equation (8):

$$\frac{\partial Pr(D(s, \ell, \bar{a}(\ell) = 1)|\boldsymbol{Z})}{\partial Pr(D(s, \ell - 1, \bar{a}(\ell - 1) = 1)|\boldsymbol{Z})} \tag{9}$$

component by component.[29]

Table 15 presents estimates of $Q(s, a+j, a)$ from level ℓ to level $\ell+1$. We compute this transmission term $Q(s, a+j, a)$ at endpoints at each level. There are significant transmission effects of accumulated capital across levels, which are consistent across all UHP Language levels.[30] The transmission effects are predicted to be positive, and they are. Tables 15-17 present the estimates of $Q(s, a+j, a)$ from level ℓ to level $\ell + 1$ using the sample of the children with *full exposure* of all UHP language levels

[27] The technology relies on the Markovian structure, although the capitals in different periods may be in different units and may represent fundamentally different skills.

[28] The logit is $\Pr(D = 1|\boldsymbol{z}) = \frac{\exp(\gamma' z)}{1+\exp(\gamma' z)}$.

[29] We estimate each probability by logit and approximate (9) by finite differences $\frac{\Delta Pr(D(s,\ell,\bar{a}(\ell))=1),\boldsymbol{Z})}{\Delta Pr(D(s,\ell-1,\bar{a}(\ell-1))=1),\boldsymbol{Z})}$ and form the sample mean of the differences.

[30] We have comparable results for other skills. See Appendix J where comparable results are found.

and the ones who enrolled in the program at UHP Language level 6 and skipped all prior levels respectively. We find that there are significant effects of the transmission of accumulated capital from previous levels, which are generally consistent across all UHP Language levels for both groups of children. The estimated effects are comparable across strata and are statistically significant.

We find that poorly-educated grandparents and the high percentage of left behind children in villages (peer effects) have significant negative impacts on children's transmission of skills. Children living in poor peer environments transmit skills less effectively.

In Appendix I, we estimate a non-diagonal $\boldsymbol{Q}$ and allow skills to cross-fertilize. As noted in Section II.3, the sampling protocol limits our ability to conduct a fully satisfactory analysis. Table I.4 presents the estimated language skill transmission considering $\boldsymbol{K}$ as a vector by including the lagged levels of other skills, but the measures of skills are not perfectly aligned. Some observations are sampled mid-interval in levels, and others are just starting. Compared to the results in Table 15, language skill still has significantly positive transmission except for level 8, where only cognitive skills transmit to language in level 8. In addition, the design of the curriculum is such that cognitive and fine motor skills are taught intensely only after language level 6, so we cannot analyze cross-effects for language levels before that.

We conduct the same exercise in Appendix I on cognitive and fine motor skills' transmission and find similar results. For cognitive skills, it is statistically significant for all levels. For fine motor, it is significant for 5 of 6 levels. These results indicate cross effects, but the data are not well-suited to secure sharp results on this issue.

Table 15: Language Skill Transmission $Q(s, \ell, a(\ell), \boldsymbol{Z})$ by levels

	UHP Language Levels ℓ								
$\Pr(D(s, \ell, \bar{a}(\ell, s))Z) = 1$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	$\ell = 7$	$\ell = 8$	$\ell = 9$	$\ell = 10$	$\ell = 11$
$Pr(D(s, \ell - 1, \bar{a}(\ell - 1, s)) = 1 \vert Z)$	0.826***	0.890***	0.905***	0.652***	0.700***	0.602***	0.785***	0.652***	0.904***
	(0.073)	(0.102)	(0.131)	(0.117)	(0.084)	(0.128)	(0.113)	(0.074)	(0.120)
Grandparent	-0.210	-0.173	-0.142	0.339**	-0.205*	-0.154	-0.061	-0.337***	0.111
	(0.187)	(0.135)	(0.205)	(0.170)	(0.120)	(0.140)	(0.179)	(0.123)	(0.180)
% Left-Behind in the Village	0.326	-0.202	0.013	-0.767***	0.068	-0.107	0.130	0.120	0.022
	(0.342)	(0.229)	(0.320)	(0.204)	(0.196)	(0.271)	(0.282)	(0.200)	(0.240)
Father's Years of Education	0.001	0.003	0.014*	-0.003	0.001	0.007	-0.004	0.000	0.001
	(0.003)	(0.003)	(0.008)	(0.003)	(0.002)	(0.005)	(0.003)	(0.002)	(0.002)
Mother's Years of Education	-0.002	-0.003	-0.012	0.002	0.000	-0.008	0.005*	-0.000	0.001
	(0.003)	(0.002)	(0.007)	(0.002)	(0.002)	(0.005)	(0.003)	(0.002)	(0.002)
Male	-0.089**	0.078	-0.024	-0.066**	0.070***	-0.012	-0.023	0.030	-0.017
	(0.044)	(0.052)	(0.047)	(0.032)	(0.027)	(0.041)	(0.044)	(0.042)	(0.031)
$Pr(D(s, \ell - 1, \bar{a}(\ell - 1, s)) = 1 \vert Z)$ stepdown p-value for estimated coefficients	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Notes: 1. Control variables ($\boldsymbol{Z}$) include gender, the fraction of time children were raised by grandparents throughout the intervention, parents' years of education, and the fraction of left-behind children at the village level. 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report average marginal effects from the logits. Standard errors of average marginal effects are clustered at the village level and are reported in parentheses. 3. Stepdown p-values for estimated coefficients from logit are constructed by multiple hypotheses between whether children's last task of the previous level is correct or not across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

Table 16: Language Skill Transmission $Q(s, \ell, a(\ell), \boldsymbol{Z})$ by levels Children with Full Exposure of Curriculum

	UHP Language Levels ℓ								
$\Pr(D(s, \ell, \bar{a}(\ell, s))Z) = 1$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	$\ell = 7$	$\ell = 8$	$\ell = 9$	$\ell = 10$	$\ell = 11$
$Pr(D(s, \ell - 1, \bar{a}(\ell - 1, s)) = 1 \vert Z)$	0.830***	0.934***	0.971***	0.542***	0.848***	0.457***	0.733***	0.617***	0.798***
	(0.077)	(0.097)	(0.125)	(0.134)	(0.146)	(0.142)	(0.135)	(0.164)	(0.138)
Grandparent	-0.191	-0.052	-0.040	0.416**	-0.215	-0.225	-0.275	0.051	0.259
	(0.199)	(0.140)	(0.211)	(0.176)	(0.163)	(0.158)	(0.203)	(0.201)	(0.237)
% Left-Behind in the Village	0.292	-0.248	-0.001	-0.777***	-0.052	-0.136	-0.018	-0.193	0.005
	(0.397)	(0.264)	(0.308)	(0.225)	(0.250)	(0.241)	(0.375)	(0.303)	(0.280)
Father's Years of Education	0.003	0.011**	0.013*	-0.004	0.009	0.007	-0.004	0.009	-0.002
	(0.008)	(0.005)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
Mother's Years of Education	-0.004	-0.009*	-0.009	0.005	-0.007	-0.009	0.007	-0.008	0.004
	(0.008)	(0.005)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.007)
Male	-0.085*	0.073	-0.009	-0.074*	0.068	-0.007	-0.011	-0.033	-0.014
	(0.047)	(0.052)	(0.050)	(0.043)	(0.045)	(0.061)	(0.062)	(0.061)	(0.048)
$Pr(D(s, \ell - 1, \bar{a}(\ell - 1, s)) = 1 \vert Z)$ stepdown p-value for estimated coefficients	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Notes: 1. Control variables ($\boldsymbol{Z}$) include gender, the fraction of time children were raised by grandparents throughout the intervention, parents' years of education, and the fraction of left-behind children at the village level. 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report average marginal effects from the logits. Standard errors of average marginal effects are clustered at the village level and are reported in parentheses. 3. Stepdown p-values for estimated coefficients from logit are constructed by multiple hypotheses between whether children's last task of the previous level is correct or not across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

Table 17: Language Skill Transmission $Q(s, \ell, a(\ell), \boldsymbol{Z})$ by levels
Children Enrolled at UHP Level 6

	UHP Language Levels ℓ				
$\Pr(D(s, \ell, \bar{a}(\ell, s)) = 1 \vert Z) = 1$	$\ell = 7$	$\ell = 8$	$\ell = 9$	$\ell = 10$	$\ell = 11$
$Pr(D(s, \ell-1, \bar{a}(\ell-1, s)) = 1 \vert Z)$	0.686***	0.813***	0.837***	0.672***	0.996***
	(0.103)	(0.171)	(0.135)	(0.074)	(0.138)
Grandparent	-0.254	-0.057	0.210	-0.675***	-0.032
	(0.185)	(0.224)	(0.256)	(0.142)	(0.223)
% Left-Behind in the Village	0.323	-0.073	0.193	0.506*	0.064
	(0.276)	(0.386)	(0.353)	(0.272)	(0.306)
Father's Years of Education	-0.000	0.006	-0.005	0.000	0.002
	(0.002)	(0.005)	(0.004)	(0.002)	(0.002)
Mother's Years of Education	0.002	-0.007	0.006	0.000	0.001
	(0.002)	(0.005)	(0.004)	(0.002)	(0.002)
Male	0.082*	-0.019	-0.033	0.082	-0.019
	(0.046)	(0.052)	(0.056)	(0.052)	(0.051)
$Pr(D(s, \ell-1, \bar{a}(\ell-1, s)) = 1 \vert Z)$ stepdown p-value for estimated coefficients	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Notes: 1. Control variables ($\boldsymbol{Z}$) include gender, the fraction of time children were raised by grandparents throughout the intervention, parents' years of education, and the fraction of left-behind children at the village level. 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report average marginal effects from the logits. Standard errors of average marginal effects are clustered at the village level and are reported in parentheses. 3. Stepdown p-values for estimated coefficients from logit are constructed by multiple hypotheses between whether children's last task of the previous level is correct or not across UHP levels based on Romano and Wolf (2005a,b) by 5000 times of bootstrap.

We report analogous results for other skills in Appendix J.

VII Evidence on Static Complementarity

In this section, we examine evidence on static complementarity. The same issues discussed in the previous section bedevil us here. As a result, we assume separability in $\boldsymbol{K}$. The first term of Equation (5) is static complementarity which for level ℓ under separability in $\boldsymbol{K}$, can be approximated by

$$\frac{\partial^2 F^{(\ell,s)}(K(\ell-1,s,\bar{a}(\ell-1),\boldsymbol{Z}),\boldsymbol{I}(s,a(s,\ell)))}{\partial K(\ell-1,s,\bar{a}(\ell-1),\boldsymbol{Z})\partial \boldsymbol{I}'(s,a(s,\ell)))}$$

for $\underline{a}(s,\ell) \leq a(s,\ell) \leq \bar{a}(s,\ell)$.

To examine static complementarity in learning, under separability in the technology in terms of $\boldsymbol{K}$, we employ a two-stage approach that examines how the initial stock of knowledge at the beginning of a difficulty level affects the marginal returns to investment (i.e., number of tasks children are exposed to within a level regardless of mastery). Static complementarity captures the phenomenon that those who know more learn more. If a child has a higher initial stock of knowledge at the beginning of a difficulty level, we would expect the child to have higher productivity from the same number of exposures than children with lesser initial stocks of knowledge. In our first stage we define children's initial stock of knowledge using the probability of passing the last task of the previous skill level as a measure of the stock of latent skill at the end of level $\ell-1$ (i.e., $K(\ell-1,s,\bar{a}(\ell-1),\boldsymbol{Z})$). In our second stage, we test static complementarity between the stock of latent skill at the end of level $\ell-1$ and the investment (i.e., the exposure of lessons at level ℓ (i.e., $M(\ell)$)). Equation (10) describes the model we test for static complementarity for level ℓ. The logic is to shift the ℓ index of Equation (10) by one unit, and examine the average effect of increasing on lesson (i.e., $M(\ell)$) on latent skill at level ℓ.

$$\begin{aligned} Pr(D(s,\ell,\bar{a}(\ell,s))=1) = \Phi(&M(\ell), \Pr(D(s,\ell-1,\bar{a}(\ell-1,s))=1), \\ &M(\ell)\times \Pr(D(s,\ell-1,\bar{a}(\ell-1,s))=1), \boldsymbol{Z}). \end{aligned} \tag{10}$$

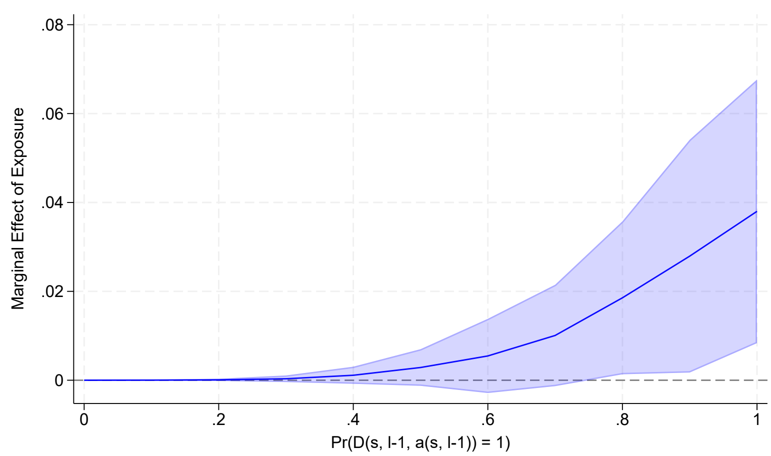
Specifically,

$$Pr(D(s,\ell,\bar{a}(\ell,s)) = 1|M(\ell),\hat{P},\boldsymbol{Z}) = \frac{1}{1+e^{-(\beta_0+\beta_1 M(\ell)+\beta_2\hat{P}+\beta_3 M(\ell)\times\hat{P}+\boldsymbol{Z}'\boldsymbol{\gamma})}} \tag{11}$$

where $D(s,\ell,\bar{a}(\ell,s))$ is an indicator of whether the child passes the last three tasks at level ℓ, $M(\ell)$ is the number of task taught at level ℓ, $\hat{P}$ is the predicted probability of passing the last task at the previous level $\ell-1$. Given the estimates from Equation (11), we calculate the average marginal effect of $M(\ell)$ at different values of stock of $K(\ell-1,s,\bar{a}(\ell-1),\boldsymbol{Z})$ (i.e., $\hat{P}$).

We present evidence of static complementarity level by level by plotting the marginal effect of exposure by initial stock of knowledge. We compute the average marginal effect of exposure using a logit model at each level and show how it varies with different levels of initial stock of knowledge, which we previously defined.

Figure 7: Static Complementarity: Average Marginal Effects of Investment to Initial Stock of Knowledge (Language Level 7)



Notes: 1. 90% confidence intervals calculated with standard errors clustered at the village level.

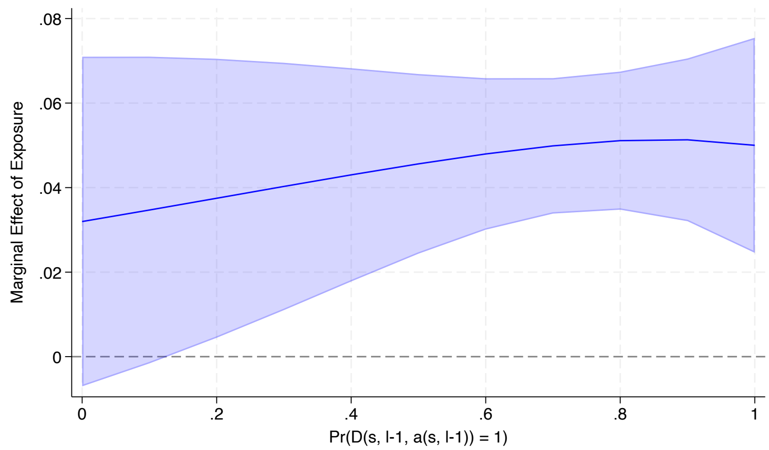
Figure 7 shows UHP Language level 7, where a positive slope indicates static complementarity - that is, children with a higher initial stock of knowledge benefit more from additional exposure. A similar pattern appears for other levels and other skills. The slopes are positive, but not all are point-wise statistically significant from zero. Some plots suggest that, as children have a high initial stock of knowledge, they almost always master the next level immediately. Thus, the marginal benefit of an additional task for children with a high stock of knowledge is small. See Appendix K.

Pooling across levels indicates a significantly positive slope to support the static complementarity. At the request of the editor, but contrary to our evidence that the skills are not comparable across levels of nominally the same skill, we estimate the model pooling across levels. Figure 8 shows a significantly positive slope to support the evidence of static complementarity for language skills. We conduct the same analysis for cognitive and fine motor skills in Appendix K.4.

Recall our discussion in Section IV. Under the Markovian assumption of Equation (1), the test of static complementarity conducted in this section is also a test of dynamic complementarity since impacts of previous investment are all encapsulated in the capital stock at the end of the previous level.

In Appendix I.3, we conduct a parallel analysis without invoking separability. The empirical results are not precisely determined, but are generally supportive of the findings reported here.

Figure 8: Static Complementarity: Average Marginal Effects of Investment to Initial Stock of Knowledge (Language, All Levels)



Notes: 1. 90% confidence intervals calculated with standard errors clustered at the village level.

VIII Learning Prior to Treatment

We test for growth in knowledge outside the experiment by examining performance on tests of skill by entrants exploiting variation in age of entry, both for treatment group children who enter at different ages and for control group children assessed by midline and endline Denver tests. We test for dynamic complementarity by exploiting the histories of lessons received by different cohorts of treated children.

Before children enroll in the program, maturation, imitation, and exposure to environments are sources of early childhood learning. To capture this, we examine the level of skills in the treatment group at their initial month of enrollment into the intervention. Since all treatment group children have comparable baseline characteristics before enrollment, the difference in the first month's performance reflects a pure out-of-program effect. Learning can arise from: (1) environments, peers, and parents, and/or (2) hardwired genetics (á la Chomsky, see Cook and Newson, 2014).

We study the growth of skills prior to entry in the program and the factors that promote it. They are input into $\boldsymbol{P}(a^*)$ in Equation (5) where a^* is the age of entry into the program. At enrollment, children have not experienced any program intervention. The marginal productivity of investment depends on the stock of capital acquired at the time of entry. We next discuss how to estimate the productivity of investment at enrollment.

VIII.1 Learning in the Treatment Group Prior to Entry

We analyze the factors affecting children's performance at enrollment (i.e., $\Pr(D(s,\ell,a^*) = 1|a^*, \boldsymbol{Z}(a^*)))$. Here, $D(s,\ell,a^*)$ is a measure of the ability to perform the task at entry for skill s at the enrollment age a^* for difficulty level ℓ. The conditioning set variables (i.e., $\boldsymbol{Z}$) includes gender, the fraction of time children were raised by the grandparent during the intervention, education of parents, and the percentage of left behind children in the same village.

Table 18 displays the estimated effects of out-of-program learning for language skills among all the treatment children at their enrollment ages. We find that the coefficient on the age of enrollment is positive and statistically significant effect on the first month passing rate, due to learning. Column (2) demonstrates the importance of the home environment in promoting such learning. The children who were raised by their grandparents learn less. We find comparable results for other skills (see Appendix L.1). We consistently find that less-educated grandparents and children with low-quality environments, as measured by the percentage of left-behind children in their village, have significant negative impacts on children's learning.

Table 18: Estimated Effect of Learning Outside of the Program for Treatment Group Using First Performance on Entry in the Programs (Language)

$\Pr(D(a^*, s, \ell) = 1 \mid a^*, Z(a^*))$	(1)	(2)
Enrollment Age (a^*)	0.018**	0.025***
	(0.007)	(0.007)
Grandparent		-0.121*
		(0.061)
$a^* \times$ % Left-Behind in the Village*		-0.072***
		(0.021)
% Left-Behind in the Village		0.772
		(0.465)
Father's Years of Education		-0.002
		(0.006)
Mother's Years of Education		0.008
		(0.005)
Male		-0.039
		(0.028)
UHP Level at First Month Control	X	X

Notes: *This is the interaction term between enrollment age (a^*) and % of left-behind children in the village. 1. Control variables ($\boldsymbol{Z}$) include gender, type of primary caregiver, parents' years of education, and fraction of left-behind children at the village level. 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report average marginal effects from our fitted OLS. Standard errors clustered at the village level are reported in parentheses.

The negative coefficient of $a^* \times$ % left-behind in the village suggests that children living in these poor-peer villages learn less. We report results for other skills in Appendix L.1. Since children are enrolled in the program at different ages, we can study the effects of learning outside of the program by level of entry. Table 19 reports that there are strong learning effects across all entry levels besides level 5.[31]

[31]Since all children enrollment levels are between levels 2 and 6, we can only test these entry

Table 19: Estimated Effect of Learning Outside of the Program for Treatment Group Across Different Entry Groups Using First Month's Entry Level Performance (Language)

	UHP Language Entry Levels ℓ				
$\Pr(D(a^*, s, \ell) = 1 \mid a^*, \boldsymbol{Z}(a^*))$	ℓ=2	ℓ=3	ℓ=4	ℓ=5	ℓ=6
Enrollment Age (a^*)	0.032**	0.032*	0.045***	0.018	0.024***
	(0.013)	(0.018)	(0.011)	(0.023)	(0.005)
Grandparent	-0.098	-0.113	-0.186*	-0.133	-0.173**
	(0.075)	(0.086)	(0.108)	(0.094)	(0.083)
$a^* \times$ % Left-Behind in the Village*	-0.211**	-0.163	-0.081	-0.133	-0.058
	(0.083)	(0.112)	(0.063)	(0.130)	(0.046)
% Left-Behind in the Village	2.339*	1.888	0.856	1.439	0.661
	(1.199)	(1.677)	(0.929)	(1.584)	(0.707)
Father's Years of Education	-0.016	-0.001	-0.003	-0.011	0.001
	(0.010)	(0.010)	(0.007)	(0.011)	(0.008)
Mother's Years of Education	0.019**	-0.010	0.011	0.001	0.008
	(0.008)	(0.009)	(0.008)	(0.009)	(0.008)
Male	-0.065	-0.065	0.021	-0.005	-0.021
	(0.051)	(0.064)	(0.059)	(0.056)	(0.047)
Enrollment Age (a^*) stepdown p-value	**0.007**	**0.036**	**0.001**	0.275	**0.001**

Notes: 1. Control variables ($\boldsymbol{Z}$) include gender, type of primary caregiver, parents' years of education, and fraction of left-behind children at the village level. 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report average marginal effects from our fitted OLS. Standard errors clustered at the village level are reported in parentheses. 3. Stepdown p-values are constructed by multiple hypotheses of children's monthly ages at enrollment across UHP levels based on Romano and Wolf (2005a,b) by 1000 times of bootstrap.
*This is the interaction term between enrollment age (a^*) and % of left-behind children in the village.

VIII.2 Learning in the Control Group

We also examine learning in the control group. We focus on control group children who have comparable baseline characteristics to those in the treatment group. The measures we study in this section use items in Denver assessments (evaluated at Midline and Endline of the intervention). To make the results compatible with those reported in the previous section, we only examine the Denver items assessed in the

levels 2-6.

UHP curriculum.[32] To estimate learning outside of the program, we use data on scores at midline with the same set of control variables as used in the analysis of out-of-program learning in the treatment group. We also examine the differences between the control group, Midline, and Endline Denver performance, item by item. Midline and Endline Denver were tested 12 months apart. The control children did not receive any lessons given in the treatment curriculum, so all skill growth for this group comes from learning outside the program.

Table 20 reports estimates of learning using midline data for controls. The results parallel those reported in Table 19. However, for this paper, we cannot meaningfully compare the rates of learning for the treatment and control groups as we have too few items and children at the comparable levels 3 and 5 at the entry level to obtain meaningful comparisons from both sides. We find comparable rates of learning for other skills in Appendix L.2.

Figure 9 compares performance item by item on the Denver tests for the *same* children in the control group at Midline and Endline Denver. We thus test knowledge of exactly the same items twice, but on different occasions. All language Denver items show significant improvement at Endline Denver when the children were 12 months older. We find comparable growth for other skills (see Appendix L.3). Skills *appreciate* with the passage of time and do not depreciate as in the Ben-Porath (1967) model. Distillation and consolidation of knowledge learned boost out-of-program acquisition of skills.

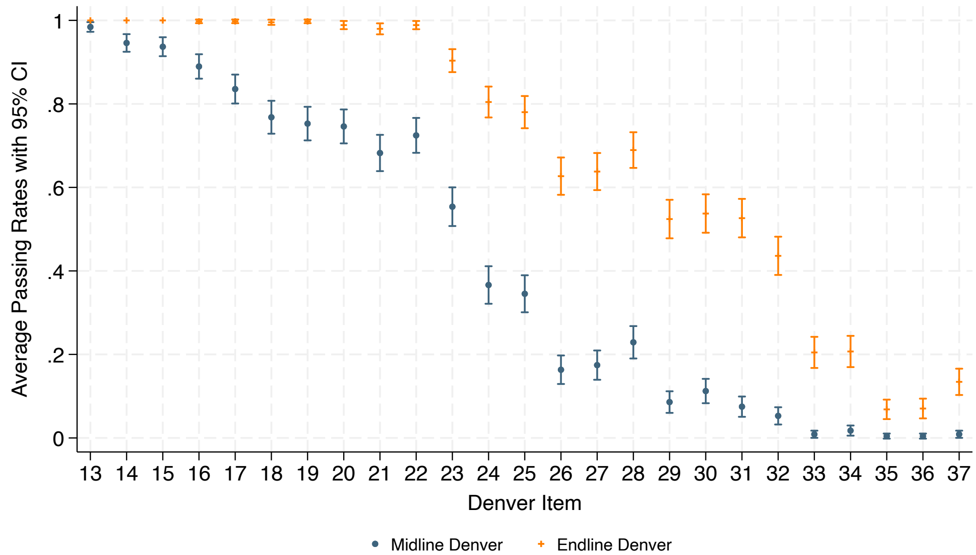
[32]Recall that we have only two such measures on both treatment and control at fixed time (not age) points. We do not use the Denver items, which are not in the UHP curriculum.

Table 20: Estimated Learning within Control Group Using Comparable Level Constructed by Midline Denver Items (Language)

	Comparable Language Level ℓ Constructed by Midline Denver Items						
	ℓ=3	ℓ=5	ℓ=7	ℓ=8	ℓ=9	ℓ=10	ℓ=11
Age at Midline	0.005***	0.018***	0.032***	0.034***	0.023***	0.079***	0.044***
	(0.001)	(0.003)	(0.003)	(0.004)	(0.004)	(0.016)	(0.008)
Grandparent	-0.013	-0.049*	-0.050	-0.056	-0.088**	-0.120	-0.060
	(0.016)	(0.027)	(0.033)	(0.039)	(0.042)	(0.074)	(0.055)
Age $\times$ % Left-Behind in the Village	-0.009	-0.013	0.002	-0.054	-0.015	0.098	-0.036
	(0.008)	(0.021)	(0.025)	(0.035)	(0.032)	(0.086)	(0.047)
% Left-Behind in the Village	0.310	0.473	0.100	1.713	0.787	-3.400	1.196
	(0.244)	(0.681)	(0.773)	(1.085)	(1.012)	(2.841)	(1.538)
Father's Years of Education	0.001	0.003	0.009***	0.010**	0.013***	-0.003	0.005
	(0.002)	(0.003)	(0.003)	(0.005)	(0.005)	(0.011)	(0.007)
Mother's Years of Education	-0.001	0.009***	0.005	0.008*	0.006	0.018**	0.012**
	(0.002)	(0.003)	(0.004)	(0.004)	(0.005)	(0.008)	(0.005)
Male	-0.007	-0.009	-0.008	-0.023	0.010	0.013	-0.015
	(0.010)	(0.018)	(0.020)	(0.027)	(0.031)	(0.061)	(0.037)
Age at Midline stepdown p-value	**0.011**	**0.002**	**0.001**	**0.002**	**0.002**	**0.005**	**0.004**

Notes: 1. Control variables ($\boldsymbol{Z}$) include gender, type of primary caregiver, parents' years of education, and fraction of left-behind children at the village level. 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report average marginal effects from our fitted OLS. Standard errors clustered at the village level are reported in parentheses. 3. Comparable levels are constructed using Denver items that are comparable to the UHP curriculum. 4. Stepdown p-values are constructed by multiple hypotheses of control group's children's monthly ages at midline Denver test across UHP levels based on Romano and Wolf (2005a,b) by 1000 times of bootstrap.

Figure 9: Learning Outside of the Program: Item by Item Comparison of Control Group Children's Performance Between Midline and Endline Denver (Language)



Notes: 1. We compare the same control children's Endline and Midline Language Denver items' performance with the exact same items that were tested twice. Endline and Midline Denver had a 12-month interval. The children in the control group did not receive any intervention throughout the program.

Appendix L.3 compares, item-by-item, Denver passing rates among the same children in the control group at Midline and Endline Denver using the exact same items as they appear in the UHP tests using paired *t-tests*. All Language Denver items show significant improvement in passing rates at Endline Denver–when the children are 12 months older. The improvement in passing rates is solely an effect of out-of-program learning. We find comparable growth in other skills (see Appendix L.3). These results illustrate a crucial aspect of early childhood learning. Although the control group did not receive any intervention investment, learning outside of the program promotes human development and is a possible contributor to "fadeout" of program treatment effects.

As a robustness check for our estimated out-of-program learning effects, children of the same age range and investment should have comparable performances. Control children whose monthly ages were between 29 and 39 months at Midline and at Endline, *respectively*, are divided into two groups with same age range but measured at different points of time: Control children who were between 29 and 39 months old at the Midline Denver in July 2016, and control children who were between 29 and 39 months old at the Endline Denver in July 2017. These two groups of control children are in the same age range at Midline/Endline Denver, respectively, but differ in age by one year. This shows that children of the same age acquire the same level of skills through exposure outside the program. See Table L.11.

IX Summary and Conclusions

This paper examines the sources of learning in a prototypical home visiting program. We find that the quality of the interaction between the home visitor and the childcare giver plays a major role in promoting the learning observed, with other interactions being less influential. We further define and estimate dynamic complementarity and the three components that determine it: critical and sensitive periods in a child's life, carryover of skills to future ages, and static complementarity. We present direct evidence on dynamic complementarity and its components using unique data from a home visiting program for young children targeted to parents in rural China.

We show how these features of dynamic complementarity affect the possibilities for remediation for children with low initial conditions. Figure 5 shows the growth in competence in a skill as a result of exposure to a program. Using a variety of different measures of ability, many children who start at the bottom of the ability distribution transition to higher ability quantiles as learning proceeds. High-ability children have persistent high performance. Low-ability children as a group proceed to higher levels of competence, but are never fully remediated. Learning rates accelerate for children without early investment when they are exposed to the program.

We also examine learning that occurs outside the intervention we study. It is a potential source causing possible "fadeout" of treatment effects after programs end, but we offer no direct evidence on this question in this paper.

Dynamic complementarity *does not* operate uniformly across skill or ability levels. It is especially strong for normal- and low-ability children, suggesting it is a powerful factor for remediation for these groups, but the results suggest a strong non-linearity. Learning outside of the program is an independent source of early childhood skill formation. It is affected by the quality of the child's environment. Growth in skills from this source is not as strong as that from the intervention we studied, but it is a source of learning that potentially accounts for fadeout in treatment effects as the control group develops skills.

For both treatment and control groups, we find *appreciation* in skills with the passage of time, and not depreciation as in conventional models. The Ben-Porath (1967) model of skill depreciation with the passage of time does not describe our data. Our evidence also suggests that new skills emerge across nominally the same skill category at different levels. "Fadeout," claimed to be found in many studies, may be a consequence of inadequate measurement of what is erroneously regarded as the same skill, but we do not develop this point in this paper.[33]

Ideally, we would decompose expression (5) into its components and estimate the proportion of dynamic complementarity at each age from each source. However, given our measures of investment, this is not feasible in this study. We leave that task for the future.

[33] See Heckman and Zhou (2026) for development of this point.

References

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

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

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

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

Ben-Porath, Y. (1967, August). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*(4, Part 1), 352–365.

Bennhoff, F. H., J. L. García, and D. E. Leaf (2024). The dynastic benefits of early-childhood education: Participant benefits and family spillovers. *Journal of Human Capital 18*(1), 44–73.

Campbell, F. A. and C. T. Ramey (1994, April). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development 65*(2), 684–698.

Caucutt, E. and L. J. Lochner (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy 128*(3), 1065–1147.

Chaparro, J., A. Sojourner, and M. J. Wiswall (2020). Early childhood care and cognitive development. Technical report, National Bureau of Economic Research.

China Development Research Foundation (CDRF) (2015-2018). China rural education and child health (china REACH) [dataset].

Cook, V. J. and M. Newson (2014). *Chomsky's Universal Grammar: An Introduction* (3rd ed.). Oxford, UK: Wiley-Blackwell.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008, Fall). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Cunha, F., E. Nielsen, and B. Williams (2021). The econometrics of early childhood human capital and investments. *Annual Review of Economics 13*(1), 487–513.

Del Boca, D., C. Flinn, E. Verriest, and M. Wiswall (2025). Parenting with patience: Parental incentives and child development. *Journal of Political Economy*, Forthcoming.

Del Boca, D., C. J. Flinn, and M. Wiswall (2014). Household choices and child development. *Review of Economic Studies 81*(1), 137–185.

Del Bono, E., J. Kinsler, and R. Pavan (2022). Identification of dynamic latent factor models of skill formation with translog production. *Journal of Applied Econometrics 37*(6), 1256–1265.

Eshaghnia, S., J. J. Heckman, R. Landersø, and R. Qureshi (2025). Transmission of family influence. Under Revision, *Econometrica*.

Gertler, P., J. J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science 344*(6187), 998–1001.

Grantham-McGregor, S. and J. A. Smith (2016). Extending the jamaican early childhood development intervention. *Journal of Applied Research on Children: Informing Policy for Children at Risk 7*(2).

Hayek, F. A. v. (1941). *The Pure Theory of Capital*. Chicago: University of Chicago Press.

Heckman, J. J. and S. Mosso (2014, August). The economics of human development and social mobility. *Annual Review of Economics 6*(1), 689–733.

Heckman, J. J. and J. Zhou (2026). The microdynamics of early childhood learning. Forthcoming, "The Journal of Political Economy".

Lord, F. M. and M. R. Novick (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.

Lu, B., R. Greevy, X. Xu, and C. Beck (2011). Optimal nonbipartite matching and its statistical applications. *The American Statistician 65*(1), 21–30.

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

Meghir, C., O. Attanasio, P. Jervis, M. Day, P. Makkar, J. Behrman, P. Gupta, R. Pal, A. Phimister, N. Vernekar, et al. (2023). Early stimulation and enhanced preschool: a randomized trial. *Pediatrics 151*(Supplement 2).

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

Romano, J. P. and M. Wolf (2005a, March). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*(469), 94–108.

Romano, J. P. and M. Wolf (2005b). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

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

Todd, P. and K. I. Wolpin (2006). Assessing the impact of a school subsidy program in Mexico using experimental data to validate a dynamic behavioral model of child schooling. *American Economic Review 96*(5), 1384–1417.

Torgerson, W. S. (1958). Theory and methods of scaling.

Uzgiris, I. C. and J. M. Hunt (1975). *Assessment in Infancy: Ordinal Scales of Psychological Development*. University of Illinois Press.

van der Linden, W. J. (2016). *Handbook of Item Response Theory: Volume 1: Models*. CRC Press.

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

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

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