WORKING PAPER · NO. 2022-139

Comparing China REACH and the Jamaica Home Visiting Program

Jin Zhou, James J. Heckman, Bei Liu, Mai Lu, Susan M. Chang, and Sally Grantham-McGregor

OCTOBER 2022

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

COMPARING CHINA REACH AND THE JAMAICA HOME VISITING PROGRAM

Jin Zhou
James J. Heckman
Bei Liu
Mai Lu
Susan M. Chang
Sally Grantham-McGregor

October 2022

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number R37HD065072, the Institute for New Economic Thinking, and a grant from an anonymous donor. We thank the China Development Research Foundation for both financial and scholarly support.

Comparing China REACH and the Jamaica Home Visiting Program
Jin Zhou, James J. Heckman, Bei Liu, Mai Lu, Susan M. Chang, and Sally Grantham-McGregor
October 2022
JEL No. J13

ABSTRACT

This paper summarizes empirical findings from a series of recent papers studying a replication of the Jamaica Reach Up and Learn home visiting program in China, China REACH. It collects more detailed information than is available on the original program. An analysis of it facilitates investigation of the skills generated by Jamaica Reach Up and Learn. We find evidence for dynamic complementarity for medium- and low-ability children. Children who start behind only slowly catch up. Able children are an exception. Most children master its goals for skill development, but the pace of learning varies greatly among children classified by ability. The program scales well. Costs per pupil are roughly$500 (2015 USD). At the same ages, treatment effect sizes and skill growth curves are comparable across the Jamaica and China REACH interventions, despite differences in scale and differences in cultural settings. We develop a method for comparing scores on different tests by anchoring comparisons on common items.

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

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

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

Mai Lu
China Development Research Foundation
Floor 15, Tower A
Imperial International Center
No. 136, Andingmen Wai Avenue
Dongcheng District
Beijing, P.C. 100011
China
lumai@cdrf.org.cn

Susan M. Chang
Caribbean Institute for Health Research
The University of the West Indies
Kingston, Jamaica
susan.changlopez@uwimona.edu.jm

Sally Grantham-McGregor
University College London
Gower Street
London - WC1E 6BT
sallymcgregor@yahoo.com

1 Introduction

The study of early childhood investment and its consequences is an active field. Many consider early childhood investment in developing countries a valuable strategy for promoting national skill development (e.g., Britto et al., 2013; Engle et al., 2011). The search is on for effective, low-cost strategies that are adaptable to less-developed settings. Jamaica Reach Up and Learn, established some 30 years ago, is a successful home visiting program emulated worldwide (Gertler et al., 2022, 2014; Grantham-McGregor and Smith, 2016).

This paper studies a replication of the original Jamaica Reach Up and Learn program, China REACH, which was brought to scale in an impoverished region of Western China. There are more than 1500 participants compared to the roughly 100 participants in the original Jamaica study. Zhou et al. (2022) show that the program can be successfully implemented at scale. The unique implementation and data collection of China REACH make it possible to examine the mechanisms of Reach Up programs in greater depth than is possible with previous samples.

We compare the treatment effects and skill growth curves for the China REACH and Jamaica Reach Up and Learn programs at the same ages of young children. We find comparable treatment effect sizes and very similar skill growth curves during the intervention. The implementation costs of China REACH and Jamaica Reach Up and Learn are low, facilitating their application in less-developed environments. We develop and apply a method for comparing distinct tests by linking common items. We find evidence of dynamic complementarity for medium- and low-ability children.

This paper is organized as follows. Section 2 presents background on the China

REACH program. Section 3 reports treatment effects and skill growth curves for the two programs. Section 4 presents results on dynamic complementarity—whether there is any advantage to starting the curriculum early. Section 5 calculates the costs of the China REACH program and discusses the feasibility of scaling up the Jamaican prototype. Section 6 concludes.

2 China REACH Background

A growing body of research establishes the effectiveness of home visiting programs targeted to the early years in developing the skills of disadvantaged children. Promising home visiting programs are relatively cheap. They place minimal demands on the training of the visitors and on the infrastructure required to support them. The Jamaica Reach program, established some 30 years ago, is a successful prototype widely emulated around the world (Grantham-McGregor and Smith, 2016).

Little is known about the mechanisms producing its treatment effects and whether or not the program can be successfully implemented at scale. This paper addresses these two key issues. To do so, we study China REACH, a replication of the original Jamaica Reach Up and Learn program that was launched in 2015 and brought to scale in China. Like the parent Jamaica program, China REACH seeks to improve the health, cognition, and engagement of children, caregivers, and associated communities. Like the original Jamaica program, it is evaluated by a randomized control trial. Unlike the Jamaica program, China REACH does not focus exclusively on stunted children. The program we analyze closely resembles the original Jamaica program and was indeed designed by its creators.

3 Treatment Effects of the Intervention

The children in the China REACH experimental treatment group are more likely to have higher language and cognitive skills, both at midline and endline, than controls (see Table 1). On average, treated children's language scores are higher than those of children in the control group. The first row shows that at midline (about 9 months after the intervention) the language and cognitive skills for the children in the treatment group are about 0.7 standard deviations higher than the controls. At the end of the intervention, the treatment effects on language and cognitive skill have effect sizes higher than 1.1. Using comparable tests, treatment size is comparable to that of the source Jamaica Reach Up and Learn intervention (i.e., about 0.75 standardized deviations). The intervention significantly improves treated children's language and cognitive skills. Treatment effects increase for children in the treatment group who have longer exposure to home visitors (see columns (3) and (5)).

Zhou et al. (2022) develop and estimate a nonlinear factor model to assess program treatment effects. This method isolates the impact of the intervention on skills and identifies individual level latent skills for each participant. It accounts for the progression of item difficulty in the program. Table 2 presents the treatment effects for the four skill factors they identify. Except for gross motor skills, all latent skill factors in the treatment group are significantly enhanced compared to those in the control group. Figure 1A shows that the distribution of language and cognitive skills

Table 1: Treatment Effects on Standardized Scores for China REACH

Denver Tasks	(1) All	(2) All	(3) Children $\leq$ 2 Yrs at Enrollment	(4) All	(5) Children $\leq$ 2 Yrs at Enrollment
			Midline		
Language and Cognitive	0.589***	0.631***	0.674***	0.714***	0.741***
	[0.234, 0.965]	[0.237, 1.036]	[0.279, 1.067]	[0.319, 1.093]	[0.350, 1.144]
Fine Motor	0.334	0.559	0.629*	0.633*	0.703*
	[-0.140, 0.787]	[-0.032, 1.174]	[0.023, 1.324]	[0.003, 1.313]	[0.057, 1.375]
Socioemotional	0.690**	0.865***	0.624***	0.879***	0.620***
	[0.260, 1.117]	[0.421, 1.312]	[0.129, 1.118]	[0.467, 1.289]	[0.204, 1.067]
Gross Motor	-0.051	-0.004	0.054	-0.015	0.010
	[-0.598, 0.478]	[-0.564, 0.577]	[-0.514, 0.640]	[-0.567, 0.554]	[-0.559, 0.584]
			Endline		
Language and Cognitive	0.979***	0.914***	1.016***	1.036***	1.113***
	[0.585, 1.402]	[0.495, 1.347]	[0.637, 1.408]	[0.644, 1.458]	[0.723, 1.510]
Fine Motor	0.585**	0.574**	0.561**	0.676***	0.645**
	[0.006, 0.956]	[0.067, 1.091]	[0.030, 1.095]	[0.180, 1.170]	[0.139, 1.158]
Socioemotional	-0.201	-0.276	-0.167	-0.222	-0.115
	[-0.596, 0.202]	[-0.688, 0.123]	[-0.553, 0.215]	[-0.636, 0.194]	[-0.491, 0.275]
Gross Motor	0.067	0.125	0.155	0.173	0.219
	[-0.479, 0.632]	[-0.392, 0.645]	[-0.406, 0.732]	[-0.322, 0.668]	[-0.294, 0.775]
Pre-Treatment Covariates	No	No	No	Yes	Yes
IPW	No	Yes	Yes	Yes	Yes

Abbreviation: IPW, Inverse Probability Weight.

1. The 95% confidence intervals in parentheses are constructed by the wild bootstrap clustered at the village level.
2. The standardized score is estimated from the pooled control group children of the Denver test.
3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

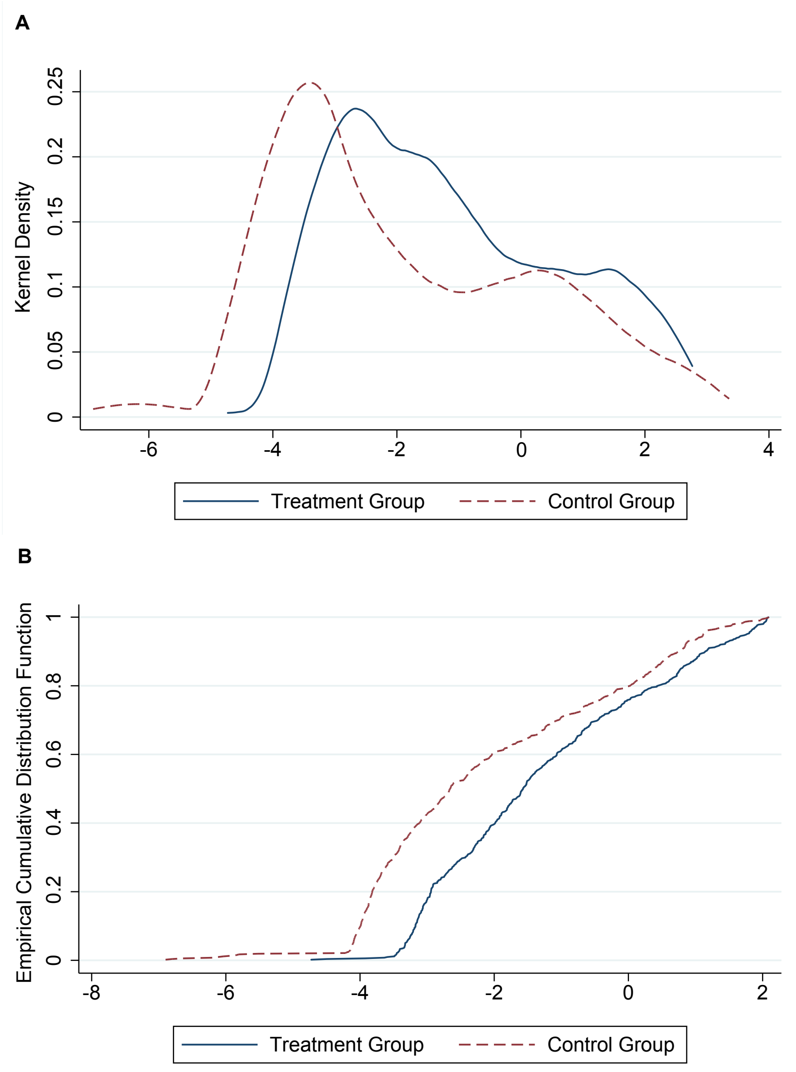
in the treatment group shifts right and has a fatter upper tail than the one in the control group. Figure 1B shows that the treated group has higher values of language skill.

Table 2: China REACH: Treatment Effects on Latent Skill Factors

	Socioemotional	Fine Motor	Language and Cognitive	Gross Motor
Treatment	0.495***	0.726***	0.753***	-0.095
	[0.208, 0.583]	[0.551, 0.899]	[0.459,1.051]	[-0.280, 0.089]

Source: Zhou et al. (2022).
1. The 95% confidence intervals in parentheses are constructed by wild bootstrap clustered at the village level.
2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1: Language and Cognitive Skill Distribution (Endline) and Dominance Curves



Source: Zhou et al. (2022).

3.1 Comparison of China REACH Treatment Effects with Those of the Original Jamaica Reach Up and Learn Program

In this section, we compare treatment effects and skill growth curves for the China REACH and Jamaica Reach Up and Learn programs. Table 3 shows the treatment effects for multiple skills. We conduct tests of equality of treatment effects for the two programs using the data available for each. We cannot reject that the treatment effect sizes are not significantly different from each other.

However, the two interventions use different tools for measuring skill development: children in China REACH were evaluated by the Denver II test, and the Griffiths test was used to evaluate children in the Jamaican program. The two tests are different. Luiz et al. (2004) compare the Denver and Griffiths tests and find that "there was a significant relationship between the overall performance of the Denver II and the Griffiths Scales. However, the Personal-Social Scale of the Denver II appeared to have items that were culturally biased. Further, the Denver II further identified a higher percentage of the sample to have abnormal or questionable protocols than the Griffiths Scales did." Elliman et al. (1985) compare both tests for premature children. Therefore, we need a more reliable method to make valid comparisons of the latent skills of the children in these two programs.

Table 3: Treatment Effects on China REACH and Jamaica Reach Up and Learn

	Socioemotional	Fine Motor	Language and Cognitive	Gross Motor
Panel A: China REACH Latent Skill Factors (After 21 Months of Intervention)				
	Socioemotional	Fine Motor	Language and Cognitive	Gross Motor
Treatment	0.40***	0.73***	0.75***	-0.10
	[0.21, 0.58]	[0.55, 0.90]	[0.46,1.05]	[-0.28, 0.09]
Panel B: Jamaica Griffiths Test (After 24 Months of Intervention)				
	Performance	Fine Motor	Hearing and Speech	Gross Motor
Treatment	0.63***	0.67***	0.50***	0.34***
	[0.30, 0.95]	[0.34, 1.00]	[0.15,0.84]	[0.01, 0.67]
p-value	0.35	0.78	0.39	0.15

Source: Zhou et al. (2022).

1. For the China REACH program, the 95% confidence intervals in brackets are constructed by wild bootstrap clustered at the village level.
2. For the Jamaica Reach Up and Learn program, the 95% confidence intervals are presented in brackets.
3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
4. The *p*-values in the last row correspond to the null of equality of treatment effects across the programs.

To conduct more reliable comparisons, we list the items in the Denver II and Griffiths tests that have the exact same content and examination criteria (see Table 4). Since these items have the same content, we use them to link the two programs.

Table 4: List of Items With Same Content in Denver and Griffiths Tests

Language	Combine words
	Say two opposites
Fine Motor	Copy circle
	Copy cross
Gross Motor	Walk alone well
	Walk backwards
	Jump off a step
	Go down stairs alone
	Throw ball

3.2 A Method for Standardizing Tests

To estimate the underlying unobserved skills across programs, we have to address the challenge that different programs use different assessment tools. We use a modified version of the Rasch model to separately estimate individual unobserved skill factors and item difficulty levels for each program. To convert the assessment outcomes from different instruments and link the different programs, we choose the items with the same content and examination criteria as anchors.

There are two types of measures in the Denver II and Griffiths tests: ordered measures $m \in M^o$ and unordered measures $m \in M^{no}$. The ordered test items are designed to reflect the fact that if children cannot perform a task with a lower requirement, they cannot achieve a harder task. For example, in the Denver II test, the items "speak one word," "speak two words," and "speak three words" are clearly stated in order.

Denoting unobserved skill by scalar $\theta_{i,t}$ for each type of measured skill, for unordered measures $m \in M^{no}$ and individual i, the latent skill $\theta_{i,t}$ is assumed to generate latent index $y^*_{m_{i,t}}$ as follows:

$$y^*_{m_{i,t}} = \beta_m + \alpha_m \theta_{i,t} + \varepsilon_{m_{i,t}}, \tag{1}$$

$$y_{m_{i,t}} = \begin{cases} 1, & y^*_{m_{i,t}} > 0. \\ 0, & y^*_{m_{i,t}} \leq 0. \end{cases}$$

For ordered measures $m_j \in M^o_g$, $g \in \{1, \ldots, G\}$, $j \in \{1, \ldots, J\}$, and

$M^o = \{M_1^o, \ldots, M_G^o\}$,

$$y^*_{m^g_{j,i,t}} = \beta_{m^g_j} + \alpha_{m^g}\theta_{i,t} + \varepsilon_{m^g_{j,i,t}}, \tag{2}$$

$$y_{m^g_{j,i,t}} = \begin{cases} 1 & y^*_{m^g_{j,i,t}} > \eta_j, \\ 0 & y^*_{m^g_{j,i,t}} \leq \eta_j, \end{cases}$$

where $j \in \{1, \ldots, J\}$ and $\eta_1 < \eta_2 < \ldots < \eta_{J-1} < \eta_J$. We distinguish between ordered and unordered items because the Rasch model assumes that error terms are independent across items, which means that a child who fails an easier task has a positive probability of passing a harder task. However, this assumption does not hold for ordered items. Therefore, we model ordered items using an ordered probit model. We use a probit and ordered probit model to link all the items in both the Griffiths and Denver tests. In principle, we could control for family background in analyzing the China data, but Zhou et al. (2022) show that baseline family background does not significantly improve treatment effects on skills, and home environment measures are not available for the Jamaican intervention.

If item m_k in the Denver II test and item m_h in the Griffiths test examine the same content under the same examination criteria, we define those items as anchor items. For the anchor items, we require that the difficulty parameters be the same (i.e., $\beta_{m_k} = \beta_{m_h}$) and that the factor loadings between the two interventions be the same (i.e., $\alpha_{m_k} = \alpha_{m_h}$).

We estimate Equations (1)–(2) pooling Jamaica Reach Up and China REACH data jointly. For each item in both the Denver and Griffiths tests, we get estimates of difficulty level parameters β_m and latent factor loadings α_m. In forming our estimates, we assume that the latent factor distribution is normal and estimate the parameters of mean (μ_θ) and variance (σ_θ) for the latent factor in Equations (1)–(2). We report our model estimates in Tables 5–8.

We then use an Empirical Bayes procedure to form the empirical conditional posterior distribution (i.e., $g(\theta \mid Y, X; \beta, \alpha)$) of the latent factor as follows:

$$g(\theta \mid Y, X; \beta, \alpha) = \frac{\mu(Y \mid X, \theta; \beta, \alpha, \phi(\theta))\phi(\theta)}{\int \mu(Y \mid X, \theta; \beta, \alpha, \phi(\theta))\phi(\theta)d\theta}, \tag{3}$$

where the latent factor θ's distribution parameters (i.e., $\phi(\cdot)$) is a normal density of the latent factor formed from estimates of the latent factor's mean (μ_θ) and variance (σ_θ), α is factor loadings, and β is the parameters of difficulty level in Equations (1)–(2). $\mu(\cdot)$ is the empirical density given the estimates of the factor model (β, α, and $\phi(\cdot)$), and $\int \mu(Y \mid X, \theta; \beta, \alpha, \phi(\theta))\phi(\theta)d\theta$ is likelihood. We then calculate the empirical posterior density ($g(\cdot)$) by Equation (3). The predicted individual latent factors are calculated by $\hat{\theta} = \int \theta g(\theta \mid Y, X; \beta, \alpha)d\theta$.

Table 5: Denver Language Skill Items

Items Based on Probit Model				
	β_m	s.e.	α	s.e.
Combine words	5.374	0.494	1.000	
Dada/Mama specific	8.730	1.698	0.872	0.196
Body parts 6	3.795	0.271	0.661	0.074
Name 1 color	-1.158	0.081	0.399	0.042
Count 1 block	-2.186	0.139	0.502	0.055
Understand 4 prepositions	-4.053	0.321	0.439	0.058
Opposites 2	-4.040	0.336	0.291	0.044
Item Based on Ordered Probit Model				
	Cut (β_{m^g})	s.e.	α	s.e.
3 words	-8.292	0.749	1.084	0.133
6 words	-7.233	0.671	1.084	0.133
Name 1 picture	-2.934	0.156	0.640	0.065
Name 4 pictures	0.203	0.097	0.640	0.065
Speech half understandable	-4.194	0.244	0.802	0.084
Speech all understandable	1.428	0.141	0.802	0.084
Use 2 objects	2.925	0.283	1.085	0.131
Use 3 objects	4.199	0.348	1.085	0.131
Point 2 pictures	-4.395	0.245	0.733	0.077
Point 4 pictures	-1.947	0.149	0.733	0.077
Know 2 adjectives	2.809	0.195	0.647	0.073
Know 3 adjectives	5.275	0.306	0.647	0.073

Abbreviation: s.e., standard error.

Table 6: Griffiths Language Skill Items

Items Based on Probit Model				
	β_m	s.e.	α	s.e.
Uses word combinations	5.374	0.494	1.000	
Shakes head for no	3.089	0.453	0.217	0.053
Short babbled sentences of 6+ syllables	5.496	1.451	0.383	0.134
Looks at pictures for a few seconds	3.358	0.543	0.241	0.061
Tries definitely to sing	2.799	0.368	0.201	0.045
Knows own name	5.092	1.055	0.439	0.112
Likes rhymes and jingles	2.505	0.288	0.154	0.037
Picture Vocabulary (12)	-1.395	0.185	0.320	0.046
Talks well in sentences of 6+ syllables (record)	-0.827	0.228	0.546	0.088
Names six or more objects in large picture	-1.119	0.237	0.504	0.080
Opposites 2	-4.040	0.336	0.291	0.044
Names 12 objects in large picture	-3.579	0.535	0.439	0.085
Items Based on Ordered Probit Model				
	Cut (β_{m^g})	s.e.	α	s.e.
One object in box identified	-6.862	0.444	0.733	0.077
Two objects in box identified	-6.221	0.423	0.733	0.077
Four objects in box identified	-5.188	0.390	0.733	0.077
Eight objects in box identified	-3.755	0.344	0.733	0.077
Says three clear words	-11.725	1.013	1.084	0.133
Uses 4 clear words	-10.924	0.966	1.084	0.133
Uses 5 clear words	-9.970	0.920	1.084	0.133
Uses 6 or 7 clear words	-9.516	0.896	1.084	0.133
Uses 9+ clear words	-8.513	0.836	1.084	0.133
Uses 12+ clear words	-7.609	0.778	1.084	0.133
Uses 20+ clear words	-6.351	0.691	1.084	0.133

Abbreviation: s.e., standard error.

Table 7: Griffiths Language Skill Items

Items Based on Ordered Probit Model				
	Cut (β_{m^g})	s.e.	α	s.e.
Names 4 objects in box	-1.490	0.197	0.454	0.056
Names 12 of 18 objects in box	-0.044	0.170	0.454	0.056
Names 17–18 objects in box	3.758	0.285	0.454	0.056
Repeats one six-syllable sentence	1.449	0.183	0.330	0.045
Repeats sentences of 10+ syllables	2.930	0.253	0.330	0.045
Comprehends 2+ items	2.844	0.358	0.306	0.057
Comprehends 4+ items	4.516	0.513	0.306	0.057
Picture vocabulary (1)	-2.587	0.407	0.794	0.124
Picture vocabulary (2)	-1.952	0.364	0.794	0.124
Picture vocabulary (4)	-0.880	0.302	0.794	0.124
Picture vocabulary (18+)	9.108	1.012	0.794	0.124
Uses sentences of 4+ syllables, clear speech	-1.999	0.271	0.573	0.078
Defines by use (2+)	1.103	0.228	0.573	0.078
Babbled monologue when alone	-6.829	0.811	0.596	0.093
Long Babbled sentences, some words clear	-4.503	0.609	0.596	0.093
Picture description (1+ sentences)	3.187	0.444	0.464	0.080
Picture Description (3+ sentences)	5.160	0.610	0.464	0.080
Uses 2 descriptive words	1.075	0.184	0.398	0.054
Uses 6+ descriptive words	3.416	0.300	0.398	0.054
Looks at pictures with interest	-3.578	0.351	0.385	0.052
Enjoys picture book	-2.632	0.293	0.385	0.052
Uses 2+ personal pronouns	0.382	0.195	0.510	0.072
Uses 6+ personal pronouns	3.891	0.388	0.510	0.072

Abbreviation: s.e., standard error.

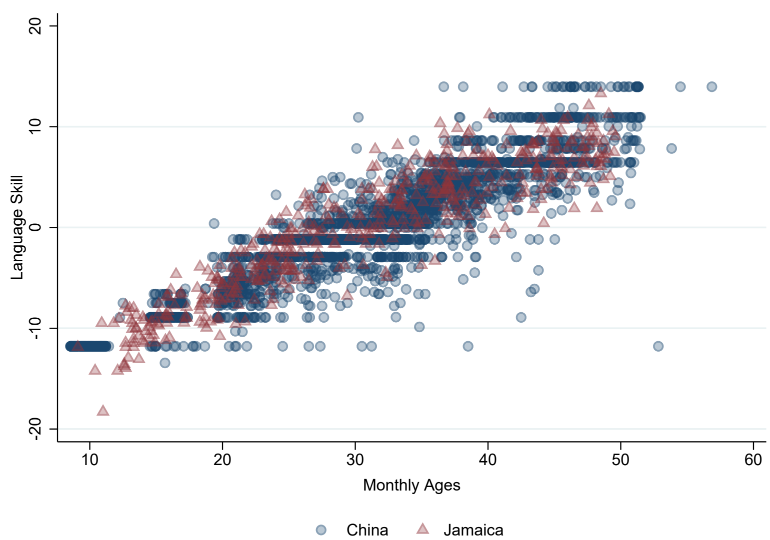
Table 8: Variance of Latent Language Skill

	Variance	s.e.
θ	39.423	7.317

Abbreviation: s.e., standard error.

Figure 2 plots the scatter of $\hat{\theta}_i$ for a model that pools language and cognitive skills for both the Jamaica Reach Up and China REACH interventions. Figure 3 plots a fitted curve based on polynomial terms of monthly ages based on $\hat{\theta}_i$.

Figure 2: Language Skill Growth Curve Comparison



We run two separate regressions: one for the treatment group and one for the control group. Then, we estimate the growth process for each program using $\hat{\theta}_i$ by treatment status as follows:

$$\hat{\theta}_i^d = \gamma_0^d + \gamma_1^g 1_{\text{China}} + \gamma_2^d \text{age}_i + \gamma_3^d \text{age}_i 1_{\text{China}} + \gamma_4^d \text{age}_i^2 + \gamma_5^d \text{age}_i^2 1_{\text{China}} + \epsilon_i^d, \quad (4)$$

where d indicates the treatment status, and 1_{China} is the indicator of whether an observation comes from a China REACH sample.

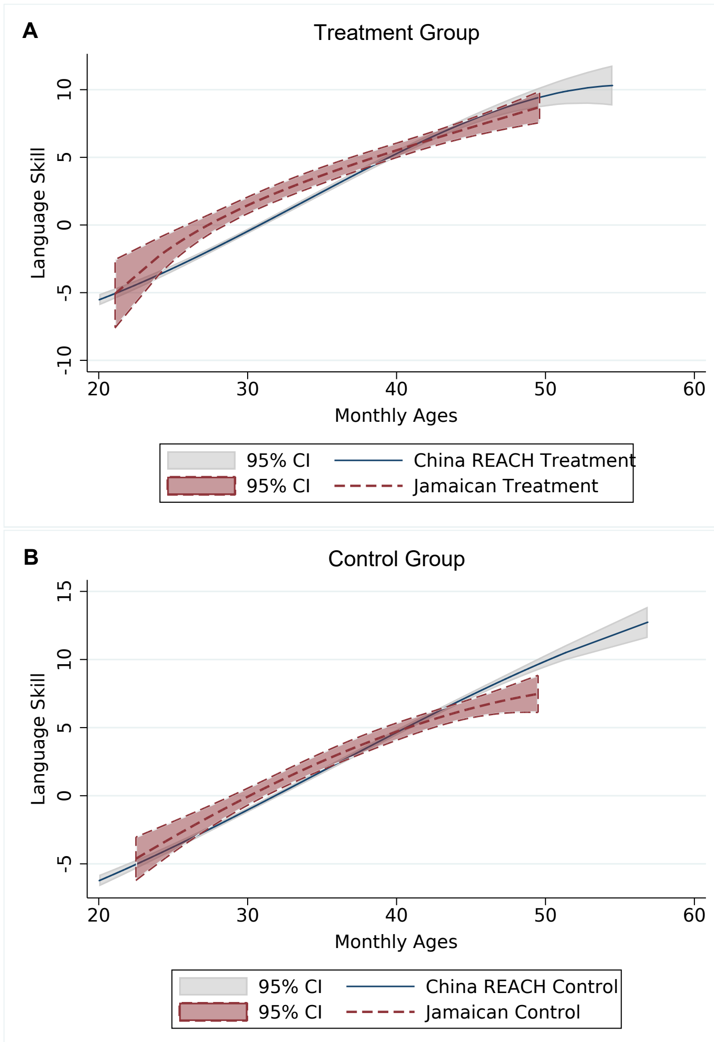
In Table 9, we provide estimates of the language skill growth curves by treatment status based on Equation (4). Our estimates imply that we cannot reject the null hypothesis that the growth curves are not significantly different between the China REACH and Jamaican interventions. For example, all of the China REACH interaction indicator coefficients are statistically insignificant. This pattern is consistent for both the treatment group and the control group, which means that the skill growth curves are not statistically significantly different between the China REACH and Jamaican interventions.

Table 9: Estimates of Language Growth Curves by Treatment Status

	Treatment	Control
Age	0.978	1.085
	[0.394, 1.563]	[0.406, 1.763]
Age$\times$ 1_{China}	-0.364	-0.545
	[-0.972, 0.243]	[-1.214, 0.125]
Age^2	-0.008	-0.009
	[-0.016, 0.002]	[-0.018,0.001]
Age^2 $\times$ 1_{China}	0.007	0.009
	[-0.002, 0.015]	[-0.001, 0.018]
Constant	-21.123	-24.703
	[-31.573, -10.672]	[-36.537,-12.869]
Constant$\times$ 1_{China}	3.264	7.410
	[-7.353, 13.883]	[-4.305, 19.125]

Figure 3 compares the language skill growth curves for China REACH and Jamaica Reach Up and Learn based on the estimates in Table 9. There is close agreement between the language skill development processes of each program. If China REACH continues on course, it will reproduce the effects of the successful Jamaica program documented in Gertler et al. (2022, 2014).

Figure 3: Language Skill Growth Curve Comparison by Treatment Status

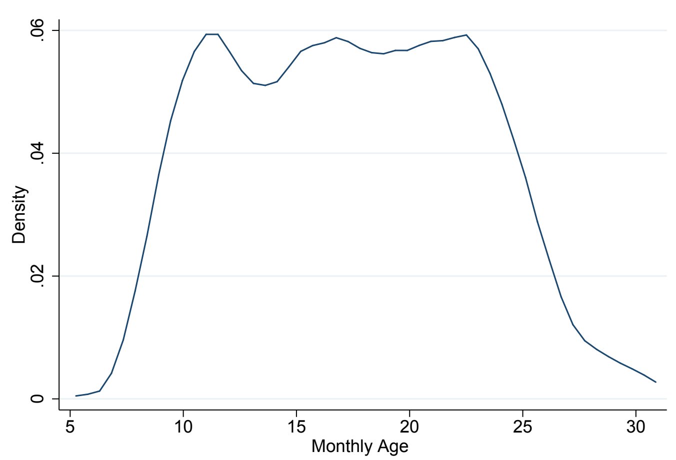


4 Dynamic Complementarity

An important question is whether investment at later ages can substitute for early childhood investment. This question is equivalent to determining the existence of dynamic complementarity (see Cunha et al., 2010): early investment improves the productivity of investment at later ages.

In the China REACH program, the uniqueness of the implementation strategy makes it possible for us to examine the existence of dynamic complementarity. Between the ages of 10 and 24 months, children enter the program more or less randomly with respect to age due to administrative constraints (see Figure 4). Since the intervention curriculum is designed based on children's weekly ages, children have the same intervention at the same weekly ages. This means that if the child is enrolled at 20 months old, he or she starts the intervention with the content for 20-month-old children and does none of the training before 20 months old in the curriculum. Similarly, if the child is enrolled at 10 months old, he or she starts with the tasks for 10-month-olds. Children who enroll at earlier ages get more investment than those who enroll at later ages.

Figure 4: The Distribution of Monthly Age When Enrolled Into the Program



Heckman and Zhou (2022b) test this hypothesis using China REACH data. Table 10 compares language passing rates at different ages for children of different ability levels who enroll early in the program with those who enroll late. In the p-value rows, they report the single null hypothesis test results at each difficulty level between the earlier enrolled group and the group enrolled at later ages.

We find a general pattern that early starters do better at the same task difficulty and child ability levels. Those who start learning earlier have persistent advantages in later life learning. Dynamic complementarity does not operate uniformly across ability groups. Medium- and low-ability children display strong dynamic complementarity effects, but high-ability children do not. We measure ability using the speed of mastery of well-defined tasks. See Heckman and Zhou (2022a). Dynamic complementarity in the China REACH program ensures that early investment improves the productivity of investment at later ages, especially for medium- and low-ability children. High-ability children catch up quickly.

5 Scaling Up: Estimating the Cost of the Program

This section discusses the per-pupil costs of the China REACH program and compares them with those of the Jamaican program. Table 11 presents the comparison between China REACH and Jamaica Reach Up and Learn. Personnel costs are the largest part for both programs. They constitute 83% for China REACH and 67% for the Jamaican program. In terms of the annual per-child cost, China REACH is about 70% of the cost of the Jamaican program. China REACH maintains a teacher-child ratio that is very close to the Jamaican program (i.e., the teacher-child ratio is about

Table 10: Language Passing Rate by Enrollment Age and Ability

	Language Difficulty Level														
Mean (Passing Rate)	7	8	9	10	11	7	8	9	10	11	7	8	9	10	11
	High-Ability					Medium-Ability					Low-Ability				
	Enroll (10–15) vs. (16–20)					Enroll (10–15) vs. (16–20)					Enroll (10–15) vs. (16–20)				
Mean (Age 10–15)	0.937	0.903	0.955	0.920	0.956	0.722	0.741	0.767	0.766	0.762	0.344	0.517	0.499	0.566	0.445
Mean (Age 16–20)	0.892	0.919	0.897	0.911	0.979	0.629	0.673	0.748	0.802	0.784	0.232	0.402	0.323	0.399	0.369
p-value	**0.080**	0.684	0.148	0.901	0.369	**0.000**	**0.005**	0.651	0.463	0.535	**0.008**	**0.021**	**0.031**	**0.084**	0.250
N	74	73	62	42	69	247	245	217	175	232	98	95	87	63	89
	Enroll (10–15) vs. (21–25)					Enroll (10–15) vs. (21–25)					Enroll (10–15) vs. (21–25)				
Mean (Age 10–15)	0.937	0.903	0.955	0.920	0.956	0.722	0.741	0.767	0.766	0.762	0.344	0.517	0.499	0.566	0.445
Mean (Age 21–25)	0.938	0.935	0.949	0.938	0.922	0.656	0.726	0.628	0.856	0.695	0.290	0.376	0.320	0.556	0.253
p-value	0.896	0.447	0.876	0.697	0.344	**0.006**	0.524	**0.004**	**0.041**	**0.065**	0.217	**0.005**	**0.030**	0.907	**0.002**
N	61	62	54	42	58	222	221	197	169	210	98	95	86	70	88
	Enroll (16–20) vs. (21–25)					Enroll (16–20) vs. (21–25)					Enroll (16–20) vs. (21–25)				
Mean (Age 16–20)	0.892	0.919	0.897	0.911	0.979	0.629	0.673	0.748	0.802	0.784	0.232	0.402	0.323	0.399	0.369
Mean (Age 21–25)	0.938	0.935	0.949	0.938	0.922	0.656	0.726	0.628	0.856	0.695	0.290	0.376	0.320	0.556	0.253
p-value	0.151	0.587	0.190	0.596	**0.028**	0.232	**0.032**	**0.010**	0.144	**0.010**	0.128	0.619	0.959	**0.061**	**0.065**
N	69	71	64	54	67	211	210	198	180	206	84	84	77	63	79

Source: Heckman and Zhou, "Nonparametric Tests of Dynamic Complementarity" (unpublished data, 2022).

1. Group (10–15) represents children whose monthly ages are between 10 and 15 at enrollment.

2. Group (16–20) represents children whose monthly ages are between 16 and 20 at enrollment.

3. Group (21–25) represents children whose monthly ages are between 21 and 25 at enrollment.

4. Fast group: the child passes the first task at over 80% of the difficulty levels, and the average passing rate at that level is greater than 80%. Normal group: the child does not pass the first task, and the passing rate is greater than 50%; or the child passes the first task, and the passing rate is between 50% and 80%. Slow group: the average passing rate is less than 50%.

5. The columns report the average passing rate from difficulty levels 7 to 11, at which all three age enrollment groups are trained during the intervention.

8 for the China REACH program and about 10 for the Jamaican program). This is promising for the scaled program.

Table 11: Program Cost per Child (Annual) Comparison Across Interventions

Category	**China REACH (Huachi)**	**Jamaica Home Visiting**
Annual Cost per Child	**527.69**	**751.60**
Fixed Cost	**91.08**	**251.47**
Expert Fee	37.54	193.10
Supplies and Facilities	53.54	58.37
Variable Cost	**436.61**	**500.13**
Personnel Cost	391.64	467.26
Toy-Making and Relevant	44.97	32.87
Teacher/Child Ratio	$93/718 \simeq 1/8$	$6/63 \simeq 1/10$

Source: China REACH cost data are collected by the program. The Jamaican program's costs are based on interviews with the original home visiting program members and the expenditure statements in historical Ford Foundation grant files. The original files presented expenditures in 1988 JMD. For both programs, after adjusting for inflation and exchange rate, we report the costs in 2015 USD.

China REACH shows that the beneficial impacts from the Jamaican program can be reproduced in a program at scale at least through the early ages. Skill requirements for being a trained home visitor are low. Visitors are residents of the villages with the same (relatively low) levels of education as the other village residents. There is an ample supply of such women. An infinitely elastic supply is a good approximation for home visitors. Initial training took two weeks and was conducted by relatively few, more highly trained program teachers who generally have advanced degrees (MA). After training, while they are in the field, local supervisors regularly monitor each home visitor. There was at least monthly field supervision of each visitor in the Jamaican intervention. Weekly group meetings and monthly observer visits were conducted for the China REACH intervention.

Visits are about an hour per week. They are adapted to conditions in the village and do not require elaborate infrastructure. The county government and the county-town-village three-tier maternal and child health care system supported the management of the program in Huachi.

6 Conclusion

This paper summarizes findings from China REACH, a replication of the original Jamaica Reach Up and Learn program, which was brought to scale in an impoverished region of Western China (over 1500 participants compared to the roughly 100 participants in the original Jamaica study). Zhou et al. (2022) show that the program can be successfully implemented at scale.

We compare the treatment effects and skill growth curves of the China REACH and Jamaica Reach Up and Learn programs. We find evidence for dynamic complementarity among low- and medium-ability groups in the replication program. We investigate the mechanisms behind the original Jamaica program in Heckman and Zhou (2022b). We also quantify the evidence that higher interaction quality between the home visitor and caregiver significantly improves treated children's skill development. We develop and implement a method for comparing diverse test scores.

References

Britto, P. R., P. L. Engle, and C. M. Super (Eds.) (2013). *Handbook of early childhood development research and its impact on global policy.* Oxford University Press. Sponsored by UNICEF and SRCD.

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.

Elliman, A., E. Bryan, A. Elliman, P. Palmer, and L. Dubowitz (1985). Denver developmental screening test and preterm infants. *Archives of Disease in Childhood 60*(1), 20–24.

Engle, P. L., L. C. H. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. Cabral de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, and S. Iltus (2011, October). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet 378*(9799), 1339–1353.

Gertler, P., J. J. Heckman, R. Pinto, S. M. Chang, S. Grantham-McGregor, C. Vermeersch, S. Walker, and A. S. Wright (2022). Effect of the Jamaica early childhood stimulation intervention on labor market outcomes at age 31. NBER Working Paper 29292. Under Revision.

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).

Heckman, J. and J. Zhou (2022a, April). Measuring knowledge. Working Paper 29990, National Bureau of Economic Research.

Heckman, J. and J. Zhou (2022b). Nonparametric tests of dynamic complementarity. Unpublised manuscript, University of Chicago.

Luiz, D. M., C. D. Foxcroft, and A. N. Tukulu (2004). The Denver ii scales and the Griffiths scales of mental development: A correlational study. *Journal of Child and Adolescent Mental Health 16*(2), 77–81.

Zhou, J., J. Heckman, B. Liu, and M. Lu (2022). The impacts of a prototypical home visiting program on child skills. Working Paper 27356, National Bureau of Economic Research.