

NBER 工作论文系列

“慧育中国”与牙买加入户辅导项目之比较

Jin Zhou

James J. Heckman

Bei Liu

Mai Lu

Susan M. Chang

Sally Grantham-McGregor

工作论文 30529

<http://www.nber.org/papers/w30529>

美国国家经济研究局(NBER)

1050 Massachusetts Avenue

Cambridge, MA 02138

2022 年 10 月

本研究得到了美国国立卫生研究院尤尼斯·肯尼迪·施赖弗国家儿童健康与人类发展研究所（资助编号 R37HD065072）、新经济思维研究所以及一位匿名捐助者的资助。我们感谢中国发展研究基金会提供的资金及学术支持。本文所表述观点仅代表作者个人，不一定代表美国国家经济研究局的观点。

NBER 工作论文供讨论和评论之用，它们尚未经过同行评审，也未接受 NBER 董事会对其官方出版物进行的审查。

©2022 作者：Jin Zhou、James J. Heckman、Bei Liu、Mai Lu、Susan M. Chang 和 Sally Grantham-McGregor。版权所有。未经明确许可只准引用两段以内文字，且须注明全部出处，包括©声明。

“慧育中国”与牙买加入户辅导项目之比较

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

NBER 工作论文，第 30529 号

2022 年 10 月

JEL No. J13

摘要

本文总结了近期一系列有关“慧育中国”（该项目参考牙买加“Reach Up and Learn”项目入户辅导模式）研究论文的实证结果。该项目收集的信息比牙买加原项目更详细。慧育中国的研究分析，有助于了解牙买加原项目产生的技能提升的影响。我们的研究证据表明，在中低能力儿童中，存在技能的“动态互补性”。开始时落后的儿童，只能慢慢赶上来；能力强的儿童则是例外。大多数儿童都能完成既定的技能发展目标，但根据儿童能力的不同，在学习速度方面存在很大差别。慧育中国项目在其它地区较好地扩大实施。每名项目儿童的入户辅导成本约为 500 美元（按 2015 年美元汇率）。尽管两个项目的实施规模和文化环境存在差异，但对同年龄儿童的分析发现，牙买加和“慧育中国”项目的干预效果大小和儿童技能增长曲线均相似。通过将相同测评项内容进行对标，我们开发了一种对不同测评工具的分数之间开展比较的新方法。

Jin Zhou

芝加哥大学人类发展经济学研究中心

1126 East 59th Street

Chicago, IL 60637

jinzhou@uchicago.edu

Mai Lu

中国发展研究基金会

中国北京市东城区安定门外大街 136 号

皇城国际中心 A 座 15 层

100011

lumai@cdrf.org.cn

James J. Heckman

芝加哥大学人类发展经济学研究中心

1126 East 59th Street

Chicago, IL 60637

和 IZA

以及 NBER

jjh@uchicago.edu

Susan M. Chang

牙买加金斯敦西印度群岛大学加勒比健康研究所

susan.changlopez@uwimona.edu.jm

Bei Liu

中国发展研究基金会

中国北京市东城区安定门外大街 136 号

皇城国际中心 A 座 15 层

100011

liubei@cdrf.org.cn

Sally Grantham-McGregor

伦敦大学学院

Gower Street

London - WC1E 6BT

sallymcgregor@yahoo.com

1 简介

儿童早期投资及其结果已成为一个活跃的研究领域。许多人认为，发展中国家的儿童早期投资是促进国家技能发展的关键战略(如 Britto 等人, 2013 年; Engle 等人, 2011 年)。人们正在寻找适用于欠发达地区的有效策略，而且其成本要低。牙买加“Reach Up and Learn”入户辅导项目启动于约 30 年前，获得成功后在世界各地广受效仿(Gertler 等人, 2022 年、2014 年; Grantham-McGregor 和 Smith, 2016 年)。

本文研究的是牙买加模式在中国的应用——“慧育中国”项目，该项目已在中国西部的一个贫困地区推广实施。与牙买加原项目大约 100 名参与者相比，本项研究参与人员已超过 1500 名。Zhou 等人(2022 年)认为，这个项目可以大规模实施并取得成功。“慧育中国”的实施方式与数据收集方法非常独特，因此与以前样本相比，能更深入地挖掘项目背后的机制。

通过比较同年龄段儿童在“慧育中国”和牙买加“Reach Up and Learn”项目中的干预效应和技能增长曲线，我们发现两者在干预期间的干预效应量相当，技能增长曲线也极其相似。“慧育中国”和牙买加“Reach Up and Learn”项目的实施成本较低，有利于在欠发达国家推广。我们开发并应用了一种通过比较共同项目来对比不同测验评分的方法。我们发现了中低能力儿童中存在动态互补性的证据。

本文结构安排如下：第 2 部分介绍“慧育中国”项目背景；第 3 部分提供了两个项目的干预效应和技能增长曲线；第 4 部分介绍了动态互补性的结果——尽早开始课程是否有优势；第 5 部分计算了“慧育中国”项目成本，

并讨论了扩大牙买加原项目规模的可行性；第 6 部分将做一总结。

2 “慧育中国”项目背景

越来越多的研究表明，儿童早期入户辅导项目对于培养弱势儿童的技能非常有效。入户辅导项目充满希望，而且相对成本较低，它们在辅导员的专业培训和基础设施支持方面的要求最低。启动于大约 30 年前的牙买加“Reach Up and Learn”项目是一个成功典型，在世界各地广受效仿(Grantham-McGregor 和 Smith, 2016 年)。

但是，对于这种干预效应背后的机制以及该项目能否大规模实施，人们知之甚少。本文将解决这两个关键问题。为此，我们对“慧育中国”（以牙买加“Reach Up and Learn”项目为蓝本）进行了研究——该项目于 2015 年启动，后在中国广泛推广。与牙买加原项目一样，“慧育中国”旨在改善儿童、看护者和相关社区的健康、认知和参与水平。此外，它也是通过随机对照试验进行评估。与牙买加项目不同的是，“慧育中国”并不只关注发育迟缓的儿童。我们分析的这个项目与牙买加原项目非常相似，而且确实是由原班人马设计。

3 干预的干预效应

与对照组相比，“慧育中国”实验干预组的儿童在中期评估和终期评估中均显示，更有可能获得更高的语言和认知技能(参见表 1)。平均而言，干预组儿童的语言评分高于对照组儿童。从第一行可看出在中期评估(干预约 9 个月)后，干预组儿童的语言和认知技能比对照组高出大约 0.7 个标准差。干预结束时，其对语言和认知技能的干预效应超过了 1.1。通过比对试验，我们发

现干预规模与牙买加“Reach Up and Learn”干预措施相当(标准化偏差约为0.75)。干预显著提高了受干预儿童的语言和认知技能。干预组儿童接触辅导员的时间越长,干预效应越好(参见第(3)和(5)栏)。

Zhou 等人(2022 年)开发了一种非线性因子模型来评估项目干预效应。该方法分离了干预对技能的影响,并确定了每位参与者个人层面的潜在技能。它解释了项目中项目难度的变化。表 2 列出了对他们确定的四个技能因子的干预效应。除大运动技能外,干预组的所有潜在技能因子均较对照组显著增强。图 1A 显示,干预组的语言和认知技能分布向右偏移,并且比对照组的上尾更宽。从图 1B 可看出,干预组的语言技能值更高。

表 1：对“慧育中国”儿童测评标准化评分的干预效应

Denver 任务	(1) 全部	(2) 全部	(3) 参与时 ≤ 2 岁 的儿童	(4) 全部	(5) 参与时 ≤ 2 岁的 儿童
中期评估					
语言和认知	0.589*** [0.234, 0.965]	0.631*** [0.237, 1.036]	0.674*** [0.279, 1.067]	0.714*** [0.319, 1.093]	0.741*** [0.350, 1.144]
精细动作	0.334 [-0.140, 0.787]	0.559 [-0.032, 1.174]	0.629* [0.023, 1.324]	0.633* [0.003, 1.313]	0.703* [0.057, 1.375]
社交情感	0.690** [0.260, 1.117]	0.865*** [0.421, 1.312]	0.624*** [0.129, 1.118]	0.879*** [0.467, 1.289]	0.620*** [0.204, 1.067]
大运动	-0.051 [-0.598, 0.478]	-0.004 [-0.564, 0.577]	0.054 [-0.514, 0.640]	-0.015 [-0.567, 0.554]	0.010 [-0.559, 0.584]
终期评估					
语言和认知	0.979*** [0.585, 1.402]	0.914*** [0.495, 1.347]	1.016*** [0.637, 1.408]	1.036*** [0.644, 1.458]	1.113*** [0.723, 1.510]
精细动作	0.585** [0.006, 0.956]	0.574** [0.067, 1.091]	0.561** [0.030, 1.095]	0.676*** [0.180, 1.170]	0.645** [0.139, 1.158]
社交情感	-0.201 [-0.596, 0.202]	-0.276 [-0.688, 0.123]	-0.167 [-0.553, 0.215]	-0.222 [-0.636, 0.194]	-0.115 [-0.491, 0.275]
大运动	0.067 [-0.479, 0.632]	0.125 [-0.392, 0.645]	0.155 [-0.406, 0.732]	0.173 [-0.322, 0.668]	0.219 [-0.294, 0.775]
干预前协变量	否	否	否	是	是
IPW (逆概率加权)	否	是	是	是	是

1. 括号内 95%置信区间是在村一级使用原始聚类自助法进行估算。
2. 标准化评分是根据 Denver 测验中根据对照组儿童的汇总样本估算。
3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

表 2: “慧育中国”: 对潜在技能因子的干预效应

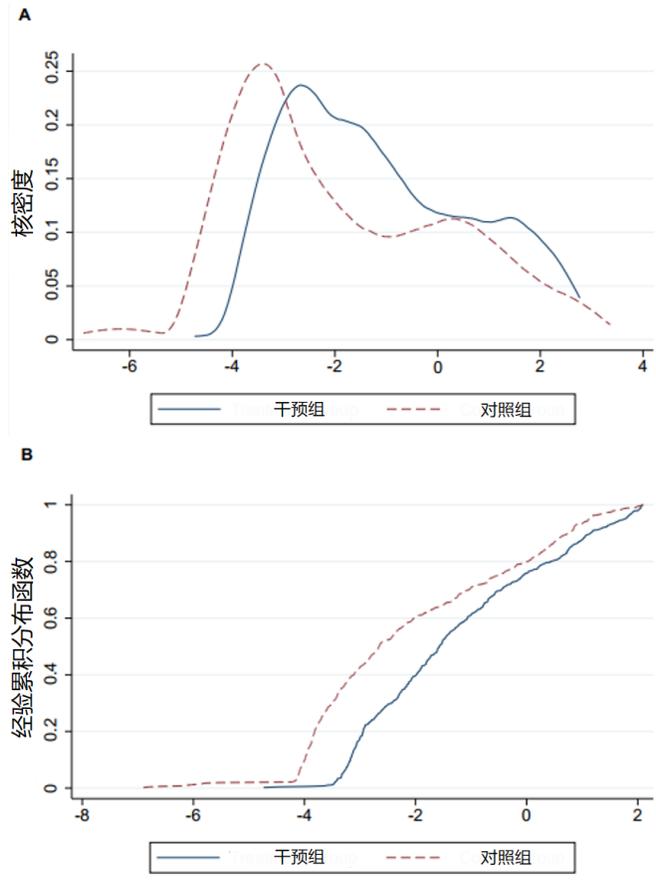
	社交情感	精细动作	语言和认知	大运动
干预组	0.495*** [0.208, 0.583]	0.726*** [0.551, 0.899]	0.753*** [0.459, 1.051]	-0.095 [-0.280, 0.089]

资料来源: Zhou 等人(2022 年)。

1. 括号内 95%置信区间是在村一级使用原始聚类自助法进行估算。

2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

图 1: 语言和认知技能分布(终期评估)和优势曲线



资料来源: Zhou 等人(2022 年)。

3.1 “慧育中国”干预效应与原牙买加“Reach Up and Learn”项目的比较

本部分中，我们将比较一下“慧育中国”和牙买加“Reach Up and Learn”项目的干预效应和技能增长曲线。表 3 展示了对多个技能的干预效应。利用两个项目的可用数据，我们对干预效应做了无差异统计检验。我们不能拒绝干预效应量彼此之间没有显著差异的原假设。

不过，这两种干预使用了不同工具来衡量技能发展：“慧育中国”借助 Denver II 测验对儿童进行评估，牙买加项目则使用了 Griffiths 测验。这两种测验截然不同。Luiz 等人(2004 年)比较了 Denver 测验和 Griffiths 测验，结果发现 Denver II 量表和 Griffiths 量表在总体表现上存在显著关系，但 Denver II 的“个人与社会”维度似乎存在文化差异。此外，与 Griffiths 量表相比，Denver II 更多诊断结果显示为“异常”或“可疑”的样本。” Elliman 等人(1985 年)比较了早产儿的两种测验情况。因此，我们需要一种更可靠的方法来对这两个项目中的儿童潜在技能进行有效比较。

表 3：“慧育中国”和牙买加“Reach Up and Learn”的干预效应

“慧育中国”潜在技能因子 (干预 21 个月后)				
	社交情感	精细动作	语言和认知	大运动
干预组	0.40*** [0.21, 0.58]	0.73*** [0.55, 0.90]	0.75*** [0.46, 1.05]	-0.10 [-0.28, 0.09]
牙买加 Griffiths 测验 (干预 24 个月后)				
	表现	精细动作	听力和语言	大运动
干预组	0.63*** [0.30, 0.95]	0.67*** [0.34, 1.00]	0.50*** [0.15, 0.84]	0.34*** [0.01, 0.67]
p 值	0.35	0.78	0.39	0.15

资料来源：Zhou 等人(2022 年)。

- 1.对于“慧育中国”项目，括号内 95%置信区间是在村一级使用原始聚类自助法进行估算。
- 2.对于牙买加“Reach Up and Learn”项目，括号内显示的是 95%置信区间。
- 3.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
- 4.最后一行的 p 值为原假设两个项目的干预效应相等的显著性检验。

为提高比较可靠性，我们列出了 Denver II 和 Griffiths 测验中，内容和检验标准完全相同的项目(参见表 4)。由于这些项目内容相同，我们便用它们来链接两个研究项目。

表 4: Denver 和 Griffiths 测验中内容相同项列表

语言	组词
	组反义词
精细动作	模仿画圈
	模仿画叉
大运动	独自行走
	后退
	跳台阶
	独自下楼
	扔球

3.2 标准化测验的方法

为估判各项目中未观察到的潜在技能，我们必须解决不同项目所使用评估工具各异的挑战。我们对 Rasch 模型做了改良，然后用它来分别估计每个项目中未观察到的技能因子和项目难度级别。为转换不同工具的评估结果并把不同项目链接在一起，我们选择了具有相同内容和检验标准的项目作为锚点。

Denver II 和 Griffiths 测验中有两类测度：有序测度 $m \in M^o$ 和无序测度

$m \in M^{no}$ 。测验项目的排序旨在反映这样一个事实：儿童如不能完成要求较低的任务，他们就无法完成较难任务。比如，在 Denver II 测验中，“会说一个词”、“会说两个词”和“会说三个词”等项目都是按序明确列出。

对于每类测查技能都用标量 θ_{it} 表示未观察到的技能，对于无序测度 $m \in M^{no}$ 和个人 i ，假设潜在技能 θ_{it} 生成潜在指数 $y_{mi,t}^*$ ，如下所示：

$$y_{mi,t}^* = \beta_m + \alpha_m \theta_{i,t} + \varepsilon_{mi,t}, \quad (1)$$

$$y_{mi,t} = \begin{cases} 1, & y_{mi,t}^* > 0. \\ 0, & y_{mi,t}^* \leq 0. \end{cases}$$

对于有序测度 $m_j \in M_g^o$, $g \in \{1, \dots, G\}$, $j \in \{1, \dots, J\}$ ，以及 $M^o = \{M_1^o, \dots, M_G^o\}$ ，有：

$$y_{m_j,i,t}^* = \beta_{m_j^g} + \alpha_{m_j^g} \theta_{i,t} + \varepsilon_{m_j,i,t}^g, \quad (2)$$

$$y_{m_j,i,t}^g = \begin{cases} 1, & y_{m_j,i,t}^* > \eta_j, \\ 0, & y_{m_j,i,t}^* \leq \eta_j, \end{cases}$$

其中 $j \in \{1, \dots, J\}$ 和 $\eta_1 < \eta_2 < \dots < \eta_{J-1} < \eta_J$ 。我们之所以区分有序项目和无序项目，是因为 Rasch 模型假设误差项在项目间是独立的，这意味着未能通过较简单任务的儿童，通过较难任务的概率为正。然而，这种假设对于有序项目并不成立，因此我们使用有序 Probit 模型对有序项目进行了建模。我们使用 Probit 和有序 Probit 模型把 Griffiths 和 Denver 测验中的所有项目链接在一起。原则上，我们在分析中国数据时可以控制家庭背景变量，但 Zhou 等人(2022 年)指出，基线家庭背景并不能显著改善对技能的干预效应，而且牙买加的干预措施也没有家庭环境测度。

如果 Denver II 测验中的 m_k 项和 Griffiths 测验中的 m_h 项在相同检验标准下检验的内容相同, 我们将这些项目定义为锚项目。对于锚项目, 我们要求难度参数相同(即 $\beta_{mk} = \beta_{mh}$), 并且两次干预之间的因子载荷相同(即 $\alpha_{mk} = \alpha_{mh}$)。

我们将牙买加 “Reach Up and Learn” 和 “慧育中国” 数据结合起来, 估算了等式(1)-(2)。对于 Denver 和 Griffiths 测验中的每个项目, 我们估算了难度级别参数 β_m 和潜在因子载荷 α_m 。在估算时, 我们假设潜在因子分布为正态, 并估计了等式(1)-(2)中潜在因子的均值(μ_θ)和方差(σ_θ)参数。表 5 至表 8 给出了我们的模型估计值。

然后, 我们使用经验贝叶斯方法来研究潜在因子的经验条件后验分布(即 $g(\theta | Y, X; \beta, \alpha)$), 如下所示:

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

其中, 潜在因子 θ 的分布参数(即 $\phi(\cdot)$)是潜在因子的标准密度, 由潜在因子均值(μ_θ)和方差(σ_θ)估计值形成, α 是因子载荷, β 是参数等式(1)-(2)中的难度级别。 $\mu(\cdot)$ 是给定因子模型(β, α 和 $\phi(\cdot)$)估计值的经验密度, $\int \mu(Y | X, \theta; \beta, \alpha, \phi(\theta))\phi(\theta)d\theta$ 是似然比。然后, 我们根据等式(3)计算出经验后验密度($g(\cdot)$)。预测的个体潜在因子这么计算: $\hat{\theta} = \int \theta g(\theta | Y, X; \beta, \alpha)d\theta$ 。

表 5: Denver 语言技能项目

基于 Probit 模型的项目				
	β_m	s.e.	α	s.e.
会组词	5.374	0.494	1.000	

会针对性叫“爸爸、妈妈”	8.730	1.698	0.872	0.196
能说出 6 个身体部位	3.795	0.271	0.661	0.074
能说出 1 种颜色	-1.158	0.081	0.399	0.042
能数出 1 个积木	-2.186	0.139	0.502	0.055
能理解 4 个介词	-4.053	0.321	0.439	0.058
能说出 2 个反义词	-4.040	0.336	0.291	0.044

基于有序 Probit 模型的项目

	Cut (β_{mg})	s.e.	α	s.e.
能说出 3 个词	-8.292	0.749	1.084	0.133
能说出 6 个词	-7.233	0.671	1.084	0.133
能说出 1 幅画的名字	-2.934	0.156	0.640	0.065
能说出 4 幅画的名字	0.203	0.097	0.640	0.065
说的话可以理解一半	-4.194	0.244	0.802	0.084
说的话完全可以理解	1.428	0.141	0.802	0.084
会使用 2 个物体	2.925	0.283	1.085	0.131
会使用 3 个物体	4.199	0.348	1.085	0.131
能指出 2 幅画	-4.395	0.245	0.733	0.077
能指出 4 幅画	-1.947	0.149	0.733	0.077
认识 2 个形容词	2.809	0.195	0.647	0.073
认识 3 个形容词	5.275	0.306	0.647	0.073

缩写: s.e. = 标准误差。

表 6: Griffiths 语言技能项目

	β_m	s.e.	α	s.e.
会用词组	5.374	0.494	1.000	
会摇头表示拒绝	3.089	0.453	0.217	0.053
能含糊说出 6 个音节以上的短句	5.496	1.451	0.383	0.134
能看几秒钟图片	3.358	0.543	0.241	0.061
会试着唱	2.799	0.368	0.201	0.045

知道自己的名字	5.092	1.055	0.439	0.112
喜欢押韵句和儿歌	2.505	0.288	0.154	0.037
能说出 12 个插图词汇	-1.395	0.185	0.320	0.046
能流利说出 6 个音节以上的句子（复述）	-0.827	0.228	0.546	0.088
能说出图中 6 个或更多的物体	-1.119	0.237	0.504	0.080
能说出 2 个反义词	-4.040	0.336	0.291	0.044
能说出图中 12 个物体	-3.579	0.535	0.439	0.085

基于 Probit 序列模型的项目

	Cut (β_{mg})	s.e.	α	s.e.
能认出框中 1 个物体	-6.862	0.444	0.733	0.077
能认出框中 2 个物体	-6.221	0.423	0.733	0.077
能认出框中 4 个物体	-5.188	0.390	0.733	0.077
能认出框中 8 个物体	-3.755	0.344	0.733	0.077
能说出 3 个词，字字清晰	-11.725	1.013	1.084	0.133
会使用 4 个词，字字清晰	-10.924	0.966	1.084	0.133
会使用 5 个词，字字清晰	-9.970	0.920	1.084	0.133
会使用 6、7 个词，字字清晰	-9.516	0.896	1.084	0.133
会使用 9+个词，字字清晰	-8.513	0.836	1.084	0.133
会使用 12+个词，字字清晰	-7.609	0.778	1.084	0.133
会使用 20+个词，字字清晰	-6.351	0.691	1.084	0.133

缩写：s.e. = 标准误差。

表 7: Griffiths 语言技能项目

基于 Probit 序列模型的项目				
	Cut (β_{mg})	s.e.	α	s.e.
能说出框中 4 个物体	-1.490	0.197	0.454	0.056
能说出框中 12 个物体（共 18 个）	-0.044	0.170	0.454	0.056
能说出框中 17-18 个物体	3.758	0.285	0.454	0.056
能重复一个 6 音节的句子	1.449	0.183	0.330	0.045
能重复 10+个音节的句子	2.930	0.253	0.330	0.045
能理解 2+个项目	2.844	0.358	0.306	0.057
能理解 4+个项目	4.516	0.513	0.306	0.057

能说出 1 个插图词汇	-2.587	0.407	0.794	0.124
能说出 2 个插图词汇	-1.952	0.364	0.794	0.124
能说出 4 个插图词汇	-0.880	0.302	0.794	0.124
能说出 18+个插图词汇	9.108	1.012	0.794	0.124
能使用 4 个音节以上的句子，发音清晰	-1.999	0.271	0.573	0.078
能在使用中说清 2+句子的含义	1.103	0.228	0.573	0.078
独自一人时呀呀学语	-6.829	0.811	0.596	0.093
能说出长而无序的句子，其中一些词清晰	-4.503	0.609	0.596	0.093
会描述图片(使用 1+个句子)	3.187	0.444	0.464	0.080
会描述图片(使用 3+个句子)	5.160	0.610	0.464	0.080
能使用 2 个描述性单词	1.075	0.184	0.398	0.054
能使用 6+个描述性单词	3.416	0.300	0.398	0.054
能饶有兴趣地观看图片	-3.578	0.351	0.385	0.052
喜欢看绘本	-2.632	0.293	0.385	0.052
能使用 2+个人称代词	0.382	0.195	0.510	0.072
能使用 6+个人称代词	3.891	0.388	0.510	0.072

缩写：s.e. = 标准误差。

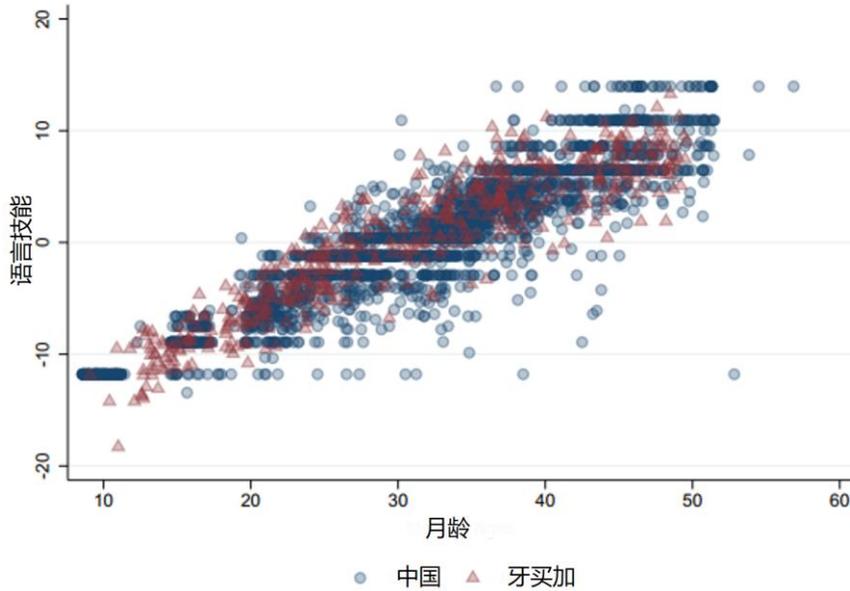
表 8：潜在语言技能方差

	方差	s.e.
θ	39.423	7.317

缩写：s.e. = 标准误差。

图 2 绘制了牙买加“Reach Up and Learn”和“慧育中国”干预措施中，语言和认知技能混合模型的散点图 $\hat{\theta}_i$ 。图 3 根据 $\hat{\theta}_i$ ，绘制了一条基于月龄多项式项的拟合曲线。

图 2：语言技能增长曲线比较



我们进行了两项单独的回归：一项针对干预组，一项针对对照组。然后我们根据干预状态，使用 $\hat{\theta}_i$ 估计了每个项目的发展过程，如下所示：

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

其中 d 表示干预状态， 1_{China} 代表观察结果是否来自“慧育中国”样本。

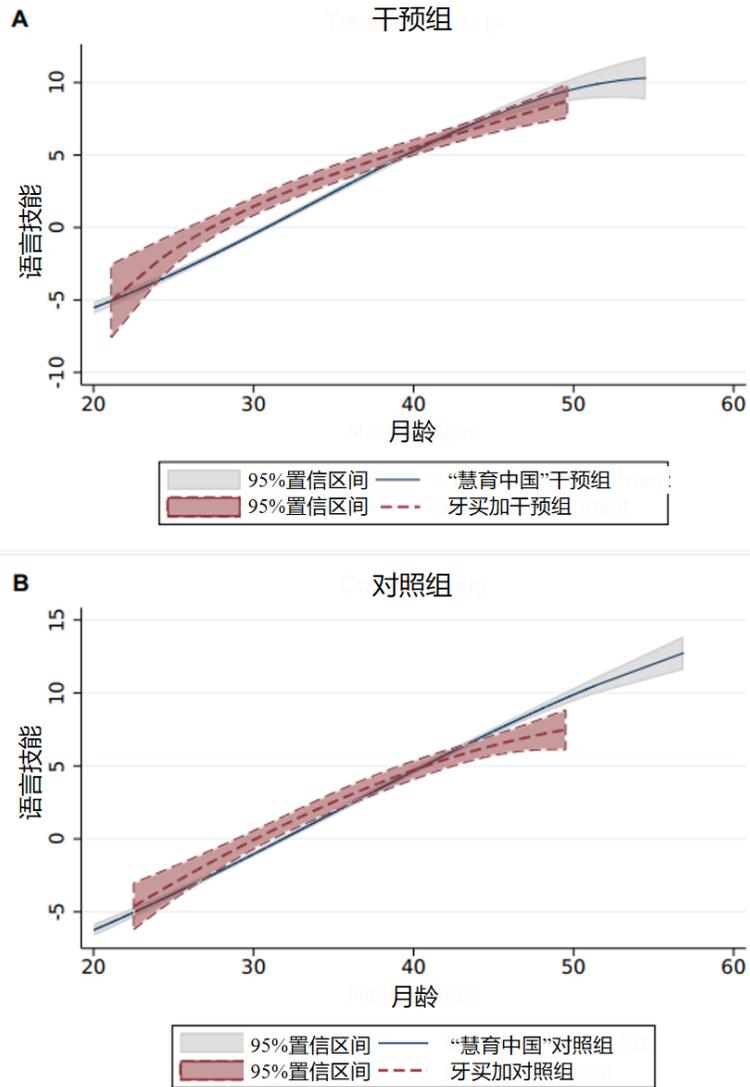
在表 9 中，我们根据等式(4)提供了按干预状态划分的语言技能增长曲线估计值。我们的估计表明我们不能拒绝原假设，即“慧育中国”和牙买加干预措施之间的增长曲线没有显著差异。比如，所有“慧育中国”交互指标系数在统计上均不显著。这种模式在干预组和对照组中一致，表明“慧育中国”和牙买加干预措施的技能增长曲线在统计学上没有明显差异。

表 9：按干预状态估计语言发展曲线

	干预组	对照组
年龄	0.978 [0.394, 1.563]	1.085 [0.406, 1.763]
Age × 1 _{China}	-0.364 [-0.972, 0.243]	-0.545 [-1.214, 0.125]
Age ²	-0.008 [-0.016, 0.002]	-0.009 [-0.018, 0.001]
Age ² × 1 _{China}	0.007 [-0.002, 0.015]	0.009 [-0.001, 0.018]
常数	-21.123 [-31.573, -10.672]	-24.703 [-36.537, -12.869]
常数 × 1 _{China}	3.264 [-7.353, 13.883]	7.410 [-4.305, 19.125]

图 3 根据表 9 中的估计值，比较了“慧育中国”和牙买加“Reach Up and Learn”的语言技能增长曲线。每个项目的语言技能发展过程都非常一致。“慧育中国”如能继续推进，必会再现牙买加项目的成功效果（Gertler 等人对此做了详细记录，2022 年、2014 年）。

图 3：两个项目干预儿童的语言技能增长曲线

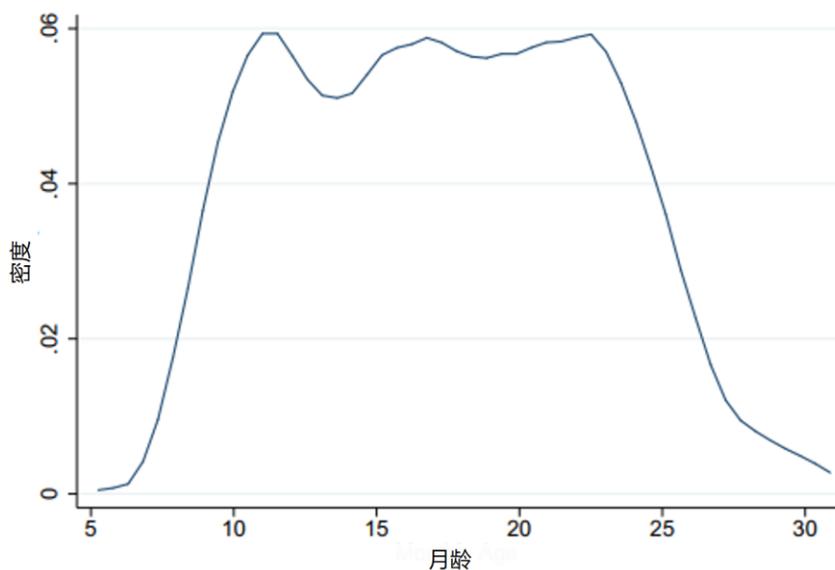


4 动态互补性

一个重要问题是，后期投资能否替代儿童早期投资。这个问题相当于确定是否存在动态互补性(参见 [Cunha 等人, 2010 年](#))：早期投资可以提高后期投资的生产率。

在“慧育中国”项目中，实施策略的独特性为我们分析是否存在动态互补性提供了机会。根据“慧育中国”项目实施方案，年龄在 10 到 24 个月之间的儿童或多或少是随机进入了该项目(参见图 4)。由于干预课程是根据儿童周龄来设计，因此周龄相同的儿童所接受的干预也一样。也就是说如果儿童是在 20 个月大时参与项目，那么他/她接受的就是针对 20 个月大儿童的干预内容，而不会接受 20 个月大之前的任何培训。同样，如果儿童是在 10 个月大时参与进来，他/她就会从 10 个月大的任务开始。较早参与的儿童比较晚参与获得的投资更多。

图 4：参与项目时儿童月龄分布



Heckman 和 Zhou(2022b)使用“慧育中国”数据检验了这一假设。表 10 根据参与项目的早晚，比较了不同能力水平的儿童在不同年龄时的语言通过率。在 p 值行中，他们给出了早参与组和晚参与组中，每个难度级别的各自原假设检验结果。

我们发现了一个普遍模式，即在任务难度和儿童能力水平相同时参与越早效果越好。开始学习较早的人在以后学习中会获得持续优势。动态互补性在不同能力群体中的运作方式并不统一。中低能力儿童表现了较强的动态互补效应，此效应在能力高的儿童则没有表现。我们根据掌握确定任务的速度来衡量能力。参见 Heckman 和 Zhou(2022a)。“慧育中国”项目具有动态互补性，因此早期投资可以提高后期投资的生产率——这在中低能力儿童身上表现尤其明显。能力高的儿童很快就能赶上。

5 扩大规模：估算项目成本

本部分讨论了“慧育中国”项目的儿童人均成本，并与牙买加项目进行了成本比较。表 11 比较了“慧育中国”与牙买加“Reach Up and Learn”项目。人员成本在这两个项目中占比最大：“慧育中国”为 83%，牙买加项目为 67%。就每人每年的成本而言，“慧育中国”约为牙买加项目的 70%。“慧育中国”的师生比（即辅导员数与儿童数配比）与牙买加项目非常接近（“慧育中国”的师生比约为 8，牙买加项目约为 10）。从这点看，扩大项目规模很有希望。

表 10: 按参与年龄和能力划分的语言通过率

均值(通过率)	语言难度														
	能力高					能力中等					能力低				
	7	8	9	10	11	7	8	9	10	11	7	8	9	10	11
	参与年龄(10 - 15) vs. (16 - 20)					参与年龄(10 - 15) vs. (16 - 20)					参与年龄(10 - 15) vs. (16 - 20)				
均值(年龄 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
均值(年龄 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 值	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
	参与年龄(10 - 15) vs. (21 - 25)					参与年龄(10 - 15) vs. (21 - 25)					参与年龄(10 - 15) vs. (21 - 25)				
均值(年龄 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
均值(年龄 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 值	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
	参与年龄(16 - 20) vs. (21 - 25)					参与年龄(16 - 20) vs. (21 - 25)					参与年龄(16 - 20) vs. (21 - 25)				
均值(年龄 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
均值(年龄 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 值	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

资料来源：Heckman 和 Zhou，“动态互补性的非参数检验”（未发表数据，2022 年）。

1.组(10-15)代表参与时年龄在 10 至 15 个月之间的儿童。

2.组(16-20)代表参与时年龄在 16 至 20 个月之间的儿童。

3.组(21-25)代表参与时年龄在 21 至 25 个月之间的儿童。

4.快速组：完成了第一个任务（难度级别在 80%以上），并且在该级别的平均通过率大于 80%。正常组：没有完成第一个任务，通过率大于 50%；或完成了第一个任务，通过率在 50%到 80%之间。慢速组：平均通过率低于 50%。

5.本表给出了难度级别在 7 至 11 时的平均通过率，所有三个年龄组都在干预期间接受了入户辅导。

表 11：比较不同干预措施下每名儿童的年成本

类别	“慧育中国” (华池县)	牙买加入户辅导项目
每名儿童的年度成本	527.69	751.60
固定成本	91.08	251.47
专家费	37.54	193.10
用品及设施	53.54	58.37
可变成本	436.61	500.13
人员成本	391.64	467.26
玩具制造及相关	44.97	32.87
师生比	93/718 ≈ 1/8	6/63 ≈ 1/10

资料来源：“慧育中国”成本数据由该项目收集。牙买加项目成本根据对原项目成员的采访以及福特基金会历史拨款档案中的支出报表计算。原始档案提供了 1988 年 JMD 的支出。对于这两个项目我们以 2015 年的美元汇率计算成本（经过通货膨胀和汇率调整后）。

“慧育中国”表明，牙买加项目的有效性可在大规模项目中复制，这至少在早期阶段是如此。经过培训，很容易培养一名辅导员，并无太高技术门槛。辅导员都是普通村民，她们的受教育程度和其他村民一样(相对较低)。这样的女性很容易找到。对辅导员来说，无限弹性供给是一个很好的近似值。上岗前的第一次培训耗时两周，由拥有高级学位、训练有素、但人数相对较少的项目培训师提供培训。培训结束后，当地督导员会定期督查每位辅导员的工作。在牙买加的干预项目中，至少每月会对辅导员进行一次现场督查。“慧育中国”干预项目每周都会召开小组形式的辅导员“集体备课”，每月督导员都会走访督查入户辅导情况。

每周每名儿童的入户辅导时间约为一小时。这项活动可因地制宜开展，不需要复杂的基础设施。县政府和县乡村三级妇幼保健机构为华池县项目管理提供了支持。

6 总结

本文总结了“慧育中国”的研究结果，该项目参考牙买加“Reach Up and Learn”项目入户辅导模式，在中国西部的一个贫困地区进行了推广(参与者超过 1500 人,而牙买加原项目的参与者约为 100 名)。Zhou 等人(2022 年)认为，这个项目可以大规模实施并取得成功。

我们比较了“慧育中国”和牙买加“Reach Up and Learn”项目的干预效应和技能增长曲线。从“慧育中国”中，我们找到了中低能力儿童群体“动态互补性”的证据。我们通过 Heckman 和 Zhou(2022b)的研究，分析了牙买加原项目背后的机制。我们还量化分析了辅导员与看护者之间，通过提高互动质量，进而显著提高干预儿童的技能发展。我们开发了一种对不同测评工具的分数之间开展比较的新方法。

参考文献

- 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 SRCDD.
- 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. Unpublished 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.