

# 教育精准扶贫中随机干预实验的中国实践与经验

史耀疆<sup>1</sup> 张林秀<sup>2</sup> 常芳<sup>1</sup> 刘涵<sup>1</sup> 唐彬<sup>1</sup> 高秋风<sup>1</sup> 关宏宇<sup>1</sup>  
聂景春<sup>1</sup> 杨洁<sup>1</sup> 白钰<sup>1</sup> 李英<sup>1</sup> 汤蕾<sup>1</sup> 岳爱<sup>1</sup> 茹彤<sup>1</sup>

(1. 陕西师范大学教育实验经济研究所, 西安 710119; 2. 联合国环境署国际生态系统管理伙伴计划、中国科学院地理科学与资源研究所生态系统网络观测与模拟重点实验室, 北京 100101)

**摘 要:** 20 世纪中后期以来, 教育研究与经济学、医学、心理学、计算机科学等学科的交叉融合不断加深, 随机干预实验作为跨学科研究方法的典型, 已逐步成为国际上制定发展政策的关键环节, 尤其是在教育和社会政策研究领域。2003 年起, 反贫困行动实验室 (The Abdul Latif Jameel Poverty Action Lab, J-PAL) 和贫困行动创新组织 (Innovations for Poverty Action, IPA) 已在全球 44 个国家开展了近 300 项教育领域的随机干预实验评估研究, 实证研究结果也被广泛应用于南亚、非洲和拉丁美洲等地区发展中国家政府的教育决策中。J-PAL 的创始人也因其对全球减贫发展的贡献荣获 2019 年诺贝尔经济学奖。虽然国际上已有大量研究验证了随机干预实验在促进教育产出、改善贫困群体收入方面的积极作用, 但中国教育精准扶贫领域的随机干预实验研究却鲜为人知。对 2007 年至 2019 年 10 月间公开发表的有关 30 多项教育精准扶贫随机干预实验研究的学术论文进行系统分析, 可以发现, 这些研究主要集中在以下四个领域: 教师与教育、信息技术与教育、营养健康与教育和儿童早期发展。这四大领域的研究以中国农村贫困地区人力资本代际陷阱为主线, 探索了低成本且可行的解决农村贫困地区教育及相关问题的有效途径, 总结了农村贫困地区教育领域随机干预实验研究的发展历史、现状和实践经验, 为农村教育实践者提供了可操作的办法, 为农村教育决策者提供了基于数据的实证支撑, 也为国家教育精准扶贫制度设计者提供了经过随机干预实验验证的有效策略。此外, 本文还对如何在开展随机干预实验的情况下使用准实验方法为教育精准扶贫政策的制定提供可靠的实证依据进行了拓展介绍。

**关键词:** 随机干预实验; 农村教育; 农村贫困地区; 教育精准扶贫

## 一、从诺贝尔经济学奖到贫困田野间的随机干预实验

### (一) 随机干预实验简介

2019 年诺贝尔经济学奖授予了三名应用实验性方案推动全球减贫发展的经济学家 Abhijit Banerjee, Esther Duflo 和 Michael Kremer, 再次将随机干预实验(Randomized Controlled Trials, RCTs)引入了大众视野。早在本世纪初, 国际上已有众多研究机构和政府重视应用随机干预实验方法进行教育、发展和公共政策项目的影响评估。2003 年, 麻省理工学院三名经济学家 Abhijit Banerjee, Esther Duflo 和 Sendhil Mullainathan 以减除全球贫困为目标, 创建了 Abdul Latif Jameel 反贫困行动实验室(The Abdul Latif Jameel Poverty Action Lab, J-PAL), 致力于推动运用包括随机干预实验在内的科学影响评估方法解决贫困问题。同年, 发展经济学家 Dean Karlan(也是 J-PAL 成员)创建了贫困行动创新组织(Innovations for Poverty Action, IPA), 持续应用随机干预实验方法为发展中国家制定公共政策提供科学的实证研究依据。十几年来, J-PAL 和 IPA 在全球 83 个国家开展了 1200 多项随机干预影响评估研究, 将影响评估广泛应用于南亚、非洲和拉丁美洲的发展中国家政府的教育决策中。

作为跨学科交叉研究方法,影响评估方法被越来越广泛地应用于经济研究、政府决策和教育改善等领域,研究者们希望以此来清楚地说明干预措施是否对最终结果产生了影响,并分析影响是如何产生的(White, 2009)。在多种影响评估方法中,随机干预实验方法被看作是识别因果关系的“黄金准则”,尤其是与传统经济学采用的非实验性方法相比(Guo et al., 2014)。随机干预实验方法早期被用来检验某种医疗方法或药物的效果,在医学等自然科学的实验中应用较多。由于能够随机分配研究对象为干预组或对照组,消除了选择性偏误,从而可将干预组和对照组在结果变量上的差异完全归因于干预因素,进而得到无偏的最接近项目“净影响”效果的估计值,近年来,随机干预实验已被广泛应用于农业经济、教育和公共政策等社会科学领域(Duflo, 2006)。除了归因外,研究者和政策制定者还可以借助“因果链”分析,理清干预或政策产生影响的作用机制,从而识别出项目或政策成功或失败的原因,为进一步完善项目设计、改善政策制定奠定基础(张林秀, 2013)。

如何开展随机干预实验呢?首先要设计随机干预实验方案,正如同样致力于推动影响评估应用的国际影响评估协会(International Initiative for Impact Evaluation, 3ie)所强调的,应在设计评估方案之初进行从投入到影响之间的整个因果链分析,再据此确定干预单位、随机方法、控制变量和实验规模等(White, 2009)。随机干预实验方案设计完成后,其执行步骤包括三个:(1)基线调查;(2)随机分配样本实施干预;(3)评估调查(张林秀, 2013)。具体操作可简述如下:

1. 开展基线调查。对所有项目研究对象开展同样的基线调查来获取样本信息,包括主要结果变量及可能影响结果变量的控制变量信息。设计调研问卷时,要涵盖因果链分析中各个环节的所有影响因素。同时也通过开展预调查不断完善调研问卷,以收集尽可能完整的信息,为下一步进行随机分配提供数据支撑。

2. 随机分配样本实施干预。基于随机分配原则,按照设计好的干预方案将所有样本分配为干预组和对照组,并在分组完成后进行平衡性检验,以确保干预组和对照组在统计意义上无显著差异。之后对干预组样本实施干预,对对照组样本则不采取任何措施。需要注意的是,干预过程中要通过多种方式实时监测干预的实施,以确保干预的依从性(Compliance)和因果链中的各项因素都得到控制,避免因果链断裂。实验过程中可能会出现干预对象不完全依从的现象,这些信息也要详细记录,为后续改善实验设计提供实践基础。

3. 开展评估调查。当干预实施到预先设计好的时长后,应对所有样本开展评估调查。评估调查内容通常与基线调查内容保持一致,即再次收集调研对象的结果变量和控制变量信息,以进行结果变量干预前和干预后的差异分析。此外,有时也需要根据实际情况再增加一部分样本的定性访谈,以帮助深入分析干预影响的因果链。根据项目干预的不同性质,评估调查可以是一次评估,也可以是多次追踪评估。在评估调查中,虽然项目实施方会采取实时监测等方式尽力追踪全部样本,但仍会遇到样本流失的问题,因此,分析结果时要计算样本流失率并进行平衡性检验,以确保干预组和对照组的流失样本不存在统计意义上的系统性差异,从而确保追踪样本分析结果无偏。

## (二) 诺贝尔奖得主应用随机干预实验推动全球减贫事业发展的案例

鉴于提升教育对促进贫困群体发展的巨大贡献,截至2016年,J-PAL和IPA的研究人员为减少全球贫困,已在44个国家开展了近300项教育评估项目,为教育扶贫提供了大量科学实证依据。

教育对提高贫困群体产出至关重要。接受更多的教育可以提升劳动生产率、增加居民收入,并相应地降低收入不平等、促进社会公平(Banerjee et al., 2013),也可以改善人们的健康状况,还可以降低犯罪率从而促进社会稳定(Lochner, 2011)。Banerjee等(2013)对1990年以来全球开展的56项教育实证研究进行回顾后发现,20世纪末至21世纪初,发展中国家的教育事业得到了快速发展,小学和中学入学率极大提升,几乎100%的公民都可以读小学。尽管各国教育事业持续快速发展,但贫困群体当前仍面临教育机会少、性别不平等、教师缺乏、学校教学质量低和学生学业表现差等问题,这些问题构成了贫困群体接受良好教育的巨大障碍(Banerjee et al., 2013)。

为应对全球教育发展面临的挑战,J-PAL 和 IPA 教育小组通过开展科学严谨的影响评估项目,验证了多种旨在提升教育产出的干预项目或试点政策的有效性,干预方式包括教师培训、教师激励、学校治理、家庭激励、信息技术和儿童早期发展干预等。这些经科学验证有效的干预项目或试点政策促进了教育产出的多方面回报,包括提升学生入学率、减少女童辍学、提升儿童读写水平等。以下通过 Duflo 等人(2012)发表在《美国经济评论》(*American Economic Review*)的文章介绍 J-PAL 如何在印度开展教师激励随机干预实验,并分析该干预对教师出勤率的提升效果和学生教育产出的影响。

1. 项目背景。在印度的偏远农村地区,非正式教学点(Non-Formal Education Centers, NFEC)是村里儿童接受教育的唯一机会。印度的一个非政府组织 Seva Mandir 在拉贾斯坦邦(Rajasthan)运营 150 个教学点,每个教学点有一名教师给 20 名左右的 7—10 岁儿童上课。NGO 工作人员发现教师缺勤率很高(平均约 35%—40%),但由于该地区地势险峻且居民居住分散,工作人员很难每日监测教师出勤情况。他们采取了一些措施降低教师缺勤率,包括不定期随访、让村民记录并报告教师出勤情况、在每月例会上跟缺勤教师谈话等,但效果并不明显,教师缺勤率依然很高。

2. 干预设计与实施。Seva Mandir 意识到之前采取的措施无法有效提高教师出勤率,在与 J-PAL 研究人员讨论后该组织决定开展一项新的教师激励随机干预实验项目。2003 年 9 月,Seva Mandir 选取了 120 所学校,随机将 60 所学校分配到干预组,剩下 60 所学校为对照组<sup>①</sup>。在干预组学校,Seva Mandir 给每名教师一台可以记录日期和时间的照相机,指导班里一名学生每天上学和放学时各给教师 and 所有学生拍一张合照。工作人员每两个月收集一次相机记录的教师出勤信息,并在教师例会上发放酬金。对干预组教师,除基本工资(1000 卢比)外,每多工作一个“有效工作日”<sup>②</sup>就可以获得 50 卢比(折合当时 1.15 美元)的奖金,而 20 天内缺勤一天则罚款 50 卢比。对照组教师依然获得与之前一样的基本工资(1000 卢比),工作人员像之前一样提醒他们每月 20 个工作日的出勤要求。

3. 数据收集。项目组通过两种方式收集教师出勤数据:每个月对所有项目实施学校随机暗访,收集干预组学校的相机拍摄记录和教师酬金发放记录。为了解教师激励干预能否提升学生教育产出,工作人员也给所有样本学生进行了三次能力测试:2003 年 8 月基线调查、2004 年 4 月(干预半年后)中期测试、2004 年 9 月(干预一年后)终期测试。基线调查分为笔试和口头考试(不会写字的儿童只参加口头考试)两种,中期测试和终期测试时所有学生都需参加笔试和口头考试。口头考试考查学生的基本数学技能,笔试则增加了数学技能的难度,也考查了学生的造句和阅读理解能力。

4. 数据分析结果。研究人员通过平衡性检验(Balance Check)验证了随机分配的有效性,通过普通最小二乘法(Ordinary Least Square, OLS)线性模型检验发现干预显著提升了教师出勤率。干预半年后,干预组教师出勤率比对照组教师显著提高了 20%;长期来看,干预 2.5 年后干预组教师出勤率仍显著高出对照组 21%。研究人员更关心的是,干预项目除提高教师出勤率外,是否提高了学生的教育产出。干预半年后,与对照组学生相比,干预组学生的语言成绩平均高出 0.16 个标准差,数学成绩高 0.15 个标准差;干预一年后干预组学生平均成绩依然比对照组学生高 0.17 个标准差。

5. 研究结论与政策含义。由于这项实验的干预方式操作简单且效果显著,截至 2009 年 10 月 NGO 仍在项目学校开展该项目。拉贾斯坦邦政府也效仿这个实验,在缺勤率高达 44% 的政府聘用护士中实施了类似的干预政策:缺勤超过 50% 工资减半,若连续缺勤则直接开除。这样的措施实行几个月后,护士缺勤率降低到了 22%,但之后随着政策“特赦”情况越来越多,既定方案没有严格执行,干预组护士缺勤率很快又回到了跟对照组一样的程度。作者分析了不同项目实施主体的不同干预效果,认为教师激励确实能有效降低教师缺勤率、提升学生教育产出,但以政府作为激励实施主体可能不易操作,从而也无法得到预期效果;若要以政府为主体开展教育领域的干预,或许聘用更多助教也可以为农村学生提供更多教育机会并提升他们的教育产出,从而减少农村学生和城市学生之间的差距。

### (三) 借鉴国际经验应用随机干预实验推进中国教育精准扶贫

2010 年中国成功进入中上等收入国家行列,伴随经济的高速发展,我国减贫事业持续推进。截至



2015年已有6亿多人口摆脱贫困,基本实现了联合国千年发展目标(习近平,2015a)。而要完全消除贫困、跨越“中等收入陷阱”迈入高收入国家行列,我国经济发展的驱动模式须从简单的依靠要素投入转向依靠技术进步和创新为主,实现这一转变的基础和前提是人力资本质量的有效提升(彭泗清等,2013;黄南,张二震,2017;谢宇等,2019)。教育作为提高人力资本质量的根本途径,是民族振兴、社会进步的重要基石,是阻断贫困代际传递的治本之策(罗仁福等,2015;全国教育大会,2018)。在发展经济和消除贫困的奋斗过程中,我国探索出了“精准扶贫”方略和“教育精准扶贫”的治本路径:扶贫必扶智、治贫先治愚(习近平,2016)。为应对当前存在的教育发展不均衡的挑战,政府相关部门提出:坚持教育公平,促进区域、城乡和各级各类教育均衡发展;要加快缩小差距,打赢教育脱贫攻坚战(李克强,2018;陈宝生,2019)。

改革开放以来,在大力发展教育事业政策的引领下,我国整体教育水平得到了极大提升,在教育方面的投入也持续增加,教育经费占GDP比例从1978年的2.1%上升到2018年的4.11%,连续7年超过4%(教育部,2019)。2010年人口普查数据显示,全国96%以上的男性和90%以上的女性均接受了不同程度的教育,教育机会的增加带来了巨大的回报(谢宇等,2019)。近年来,国家更加重视对农村贫困地区的教育投入,通过改善教育提高了贫困人口的人力资本质量,从扶贫角度看则显著提高了贫困人口的收入(Zhang & Zhao, 2006; 常芳, 2018)。尽管我国教育事业取得了巨大成就,但城乡及民族间仍存在着教育发展的不均衡,教育差距短期内会导致贫困群体劳动技能不足,长期看则会加剧未来城乡收入的不均衡,这是我国要实现消除贫困目标不容忽视的问题(Park et al., 2010; Li et al., 2015)。

2003年,J-PAL和IPA开始使用随机干预实验等科学的影响评估方法在南亚、非洲和拉丁美洲的发展中国家开展教育减贫研究,为政府教育决策提供实证依据。2005年,借鉴J-PAL和IPA的随机干预实验研究成果,我国研究者开始合作学习J-PAL和IPA的实验方法,在农村教育领域进行探索性的随机干预实验研究。2007年,国内研究团队与国外学者合作在中国开展了第一个农村教育领域的随机干预实验研究,即“农村寄宿制学校生活老师培训”随机干预实验,研究结果表明干预显著减少了寄宿学生的不良行为(Yue et al., 2014)。由于是第一次探索,这项随机干预实验存在一定的局限性,比如样本量太小(5所干预学校和5所对照学校)、统计功效(Statistical Power)<sup>③</sup>不足等。

随后的十几年里,国内研究者不断完善随机干预实验的设计和执行,合作开展了一系列采用随机干预实验方法改善农村教育的实证研究,验证了多种提高农村学生教育产出的有效手段,为促进城乡教育均衡发展、提高我国人力资本整体质量提供了可靠的实证依据。除关注学生健康,国内研究者还探索了从师资和课程设置入手改善学生学业表现的有效方式,比如,利用现代信息技术辅助教学随机干预实验弥补农村学生课后辅导资源的不足,或开展教师绩效激励随机干预实验以促进教师改善其教学行为,研究者发现这些干预均显著提高了农村学生的学业表现(Lai et al., 2016; Loyalka et al., 2019)。

除与来自埃默里大学、鲁汶大学、匹兹堡大学和斯坦福大学等国外高校的研究机构的学者合作外,国内研究团队十几年来也保持着合力探索的研究模式,这些国内合作研究团队有:安徽大学、安康学院、北京大学、河南大学、九江学院、教育部基础教育质量监测中心、兰州大学、宁夏大学、清华大学、青海民族大学、青海师范大学、陕西师范大学、四川大学、山东大学、西安交通大学、西北大学、新疆财经大学、新疆大学、香港中文大学、浙江工商大学、中国发展研究基金会、中国科学院农业政策研究中心、中国人民大学等。

本文接下来将重点介绍中国教育精准扶贫领域的随机干预实验案例,总结各团队应用随机干预实验方法在中国农村教育领域开展研究的实践经验,探讨未来如何开展随机干预实验研究以促进教育精准扶贫,以期与政府合力为教育政策制定提供基于实证的决策依据。

## 二、聚焦教育精准扶贫:随机干预实验的中国实践

人均收入最终可能达到什么样的水平,在一定程度上取决于人力资本存量。人力资本存量的增长

是经济增长的动力和主要源泉。要促进人力资本存量的增长,既要增加人力资本的数量,也要提高人力资本的质量。在决定人力资本存量水平的诸多因素当中,教育是最关键的(汪丁丁, 2010)。高水平教育可以促进经济发展,增加居民收入,改善居民生活质量(Becker, 1962; Mushkin, 1962; Grossman, 1972; Schultz, 1973; Mankiw et al., 1992; Barro, 1991; Becker, 1992; Freire-Serén, 2001)。基于2002年中国家庭收入调查(Chinese Household Income Project, CHIP)数据,陈斌开等(2010)的研究表明,教育是城乡收入差距最重要的贡献因素,其贡献程度达到43.9%。教育年限差异对城乡劳动收入差距贡献度为34.7%,教育回报差异对城乡劳动收入差距的贡献度为9.2%(陈斌开等, 2010)。

本节通过介绍过去十余年中国在农村教育和人力资本培育领域的系列研究成果,综述随机干预实验在教育精准扶贫各环节的作用,包括问题识别(贫困及其根源瞄准)、干预设计、结果评估和政策探索,希望为下一步持续消除相对贫困提供着力点和路径。

需要强调的是,教育精准扶贫不会一蹴而就,要避免孤立地看待以随机干预实验为代表的影响评估方法,而是要重视其持续解决问题的内在逻辑和价值。作为一种因果识别方法,早在1923年,随机干预实验就被用来评估肥料对于产量的影响<sup>④</sup>,并在之后被广泛应用于农村、医学和社会政策评估中。经过近100年的发展,其理论体系已趋于成熟。一个证据是,在顶级医学杂志——比如《柳叶刀》(*The Lancet*)或《英国医学期刊》(*British Medical Journal, BMJ*)——上发表采用随机干预实验方法的研究论文时,其行文写作近乎完形填空,即每一项需要说明的内容都有明确的规范写法。

在理论知识和实践技术不断成熟的情况下,随机干预实验在应用方面面临的挑战在学术领域引起了一定共鸣。2018年11月,一位哈佛大学医学院的教授作为第一作者在*BMJ*发表了一篇随机干预实验研究论文,说明了正确应用该方法的重要性。该研究使用随机干预实验评估降落伞对飞机跳伞者死亡和严重创伤的影响。行文严谨,逻辑清晰,事实清楚,符合*BMJ*发表标准。但有一个“研究不足”:被分配到干预组的样本只愿意从离地0.6米的静止飞机跳下。据此,该文得出结论:降落伞并未显著降低从飞机跳下导致的死亡和严重创伤,但未来还需研究如何将这一结论扩展到具有一定高度和速度的飞机上<sup>⑤</sup>。该文被Altmetric评选为2019年十篇最受关注论文<sup>⑥</sup>。

因此,在精准扶贫领域采用随机干预实验的难点不在方法本身,而是更多体现在如何系统地识别并回答教育领域存在的问题。首先,要通过数据深入了解一个问题,厘清问题产生的众多原因。其次,针对其根源性的主要原因设计干预,并开展随机干预实验评估。最后,根据评估结果改进干预,开始回答新问题。此环节是一个研究的结束,也是下一个评估项目的开始。回答科学问题的路径往往是:发现问题、设计干预、评估效果、发现新问题、设计新干预……。随着对问题理解的深入和对干预手段的持续研究,研究的重点从理解问题、评估有效性逐渐转变为提高干预的依从性(Compliance)<sup>⑦</sup>。

本节围绕农村教育发展的两个根源因素——教师和学生展开,具体从“教师与教育”“信息技术与教育”“营养健康与教育”和“儿童早期发展”四个方面展示这些领域的问题识别及持续解决问题的思路,旨在为今后教育精准扶贫研究提供参考经验。为充分体现国内外学者在该领域的贡献,笔者对国内外多个文献数据库进行了模糊检索和进一步精确检索,梳理了自2007年至2019年10月间公开发表论文中涉及的30多项中国教育精准扶贫领域的随机干预实验研究。检索文献数据库包括Web of science、ScienceDirect、IDEAS、Wiley、谷歌学术、中国知网和百度学术。中文检索关键词为:中国农村地区、随机干预实验、农村教育、农村教师、农村学生、农村婴幼儿、学生营养健康、信息技术与教育、儿童早期发展;英文检索关键词为:rural areas in China/ rural China, randomized intervention trial/randomized controlled trial, rural education, rural teachers, rural students/rural children, rural infants, students' nutrition and health, information technology and education, early childhood development。

本节目标是通过介绍集中关注以上四个教育研究领域的随机干预实验,来展示如何将该方法应用到精准扶贫中。由于篇幅较长,为使逻辑清晰,便于读者阅读,笔者在详细介绍各个领域的探索过程之前,先以研究过程图进行展示(图1、图2、图3、图4、图5)。

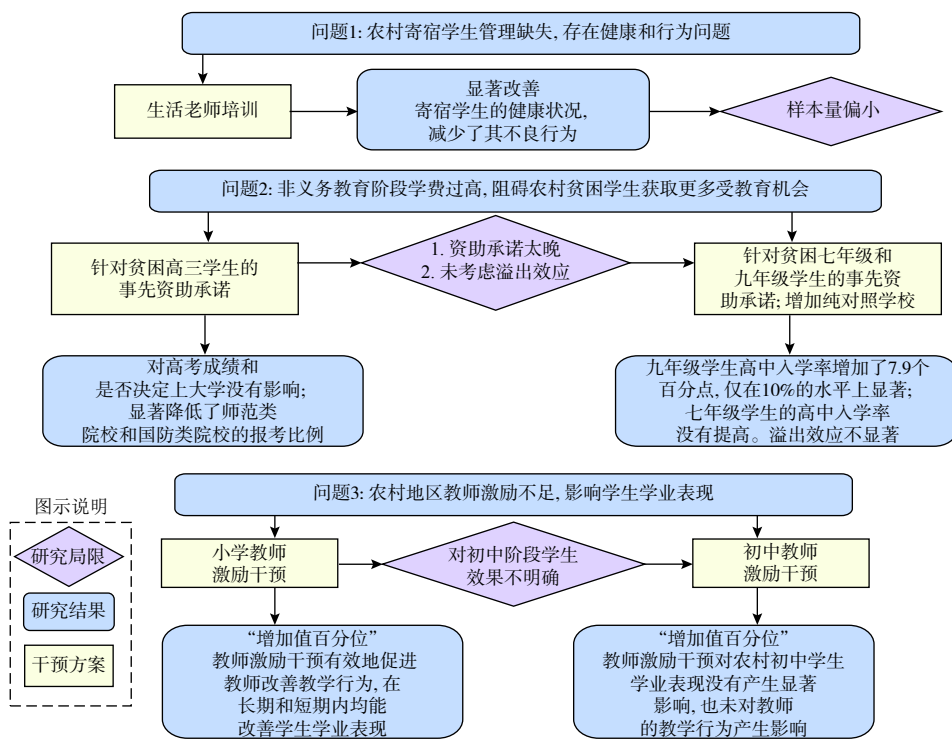


图1 “教师与教育”领域 RCT 研究过程图

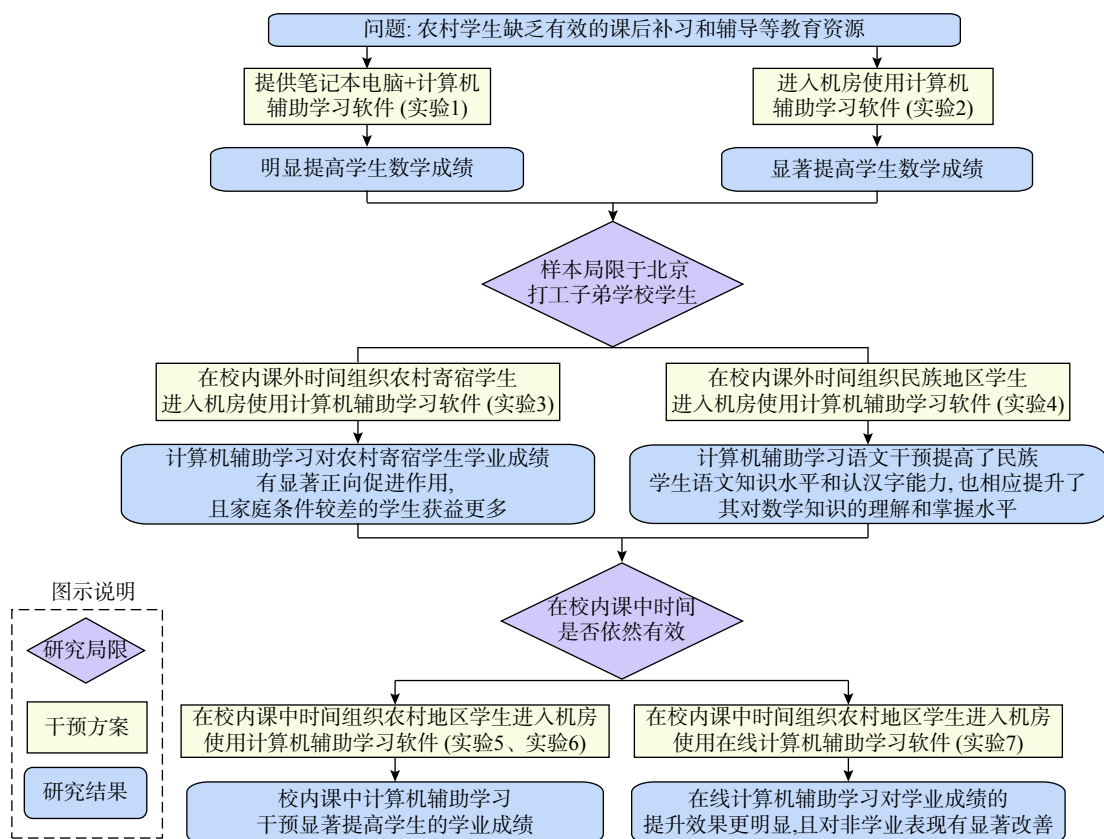


图2 “信息技术与教育”领域 RCT 研究过程图

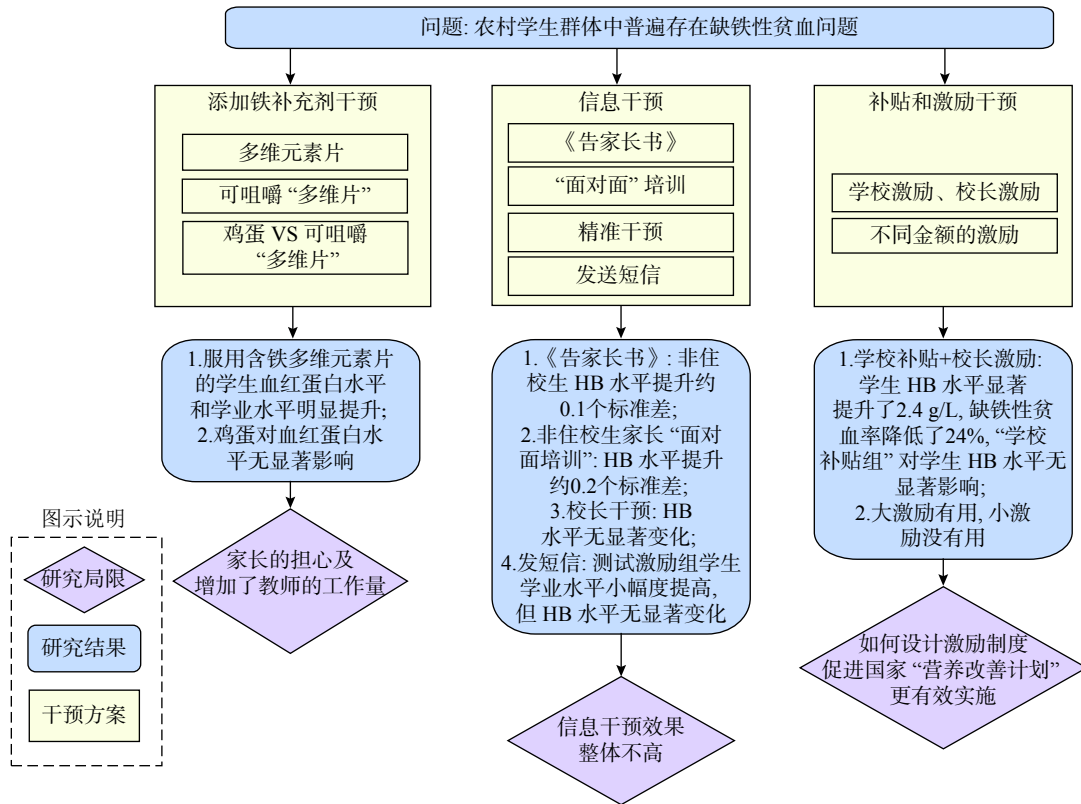


图3 “营养健康与教育”领域（缺铁性贫血）RCT 研究过程图

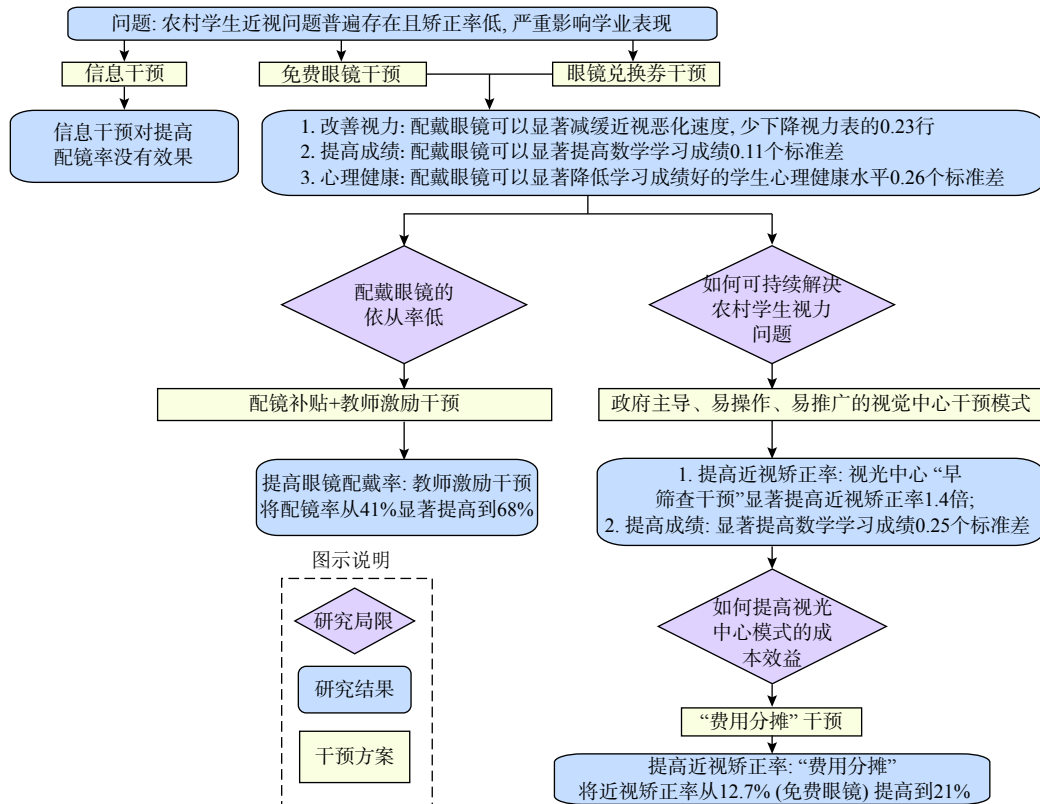


图4 “营养健康与教育”领域（近视）RCT 研究过程图



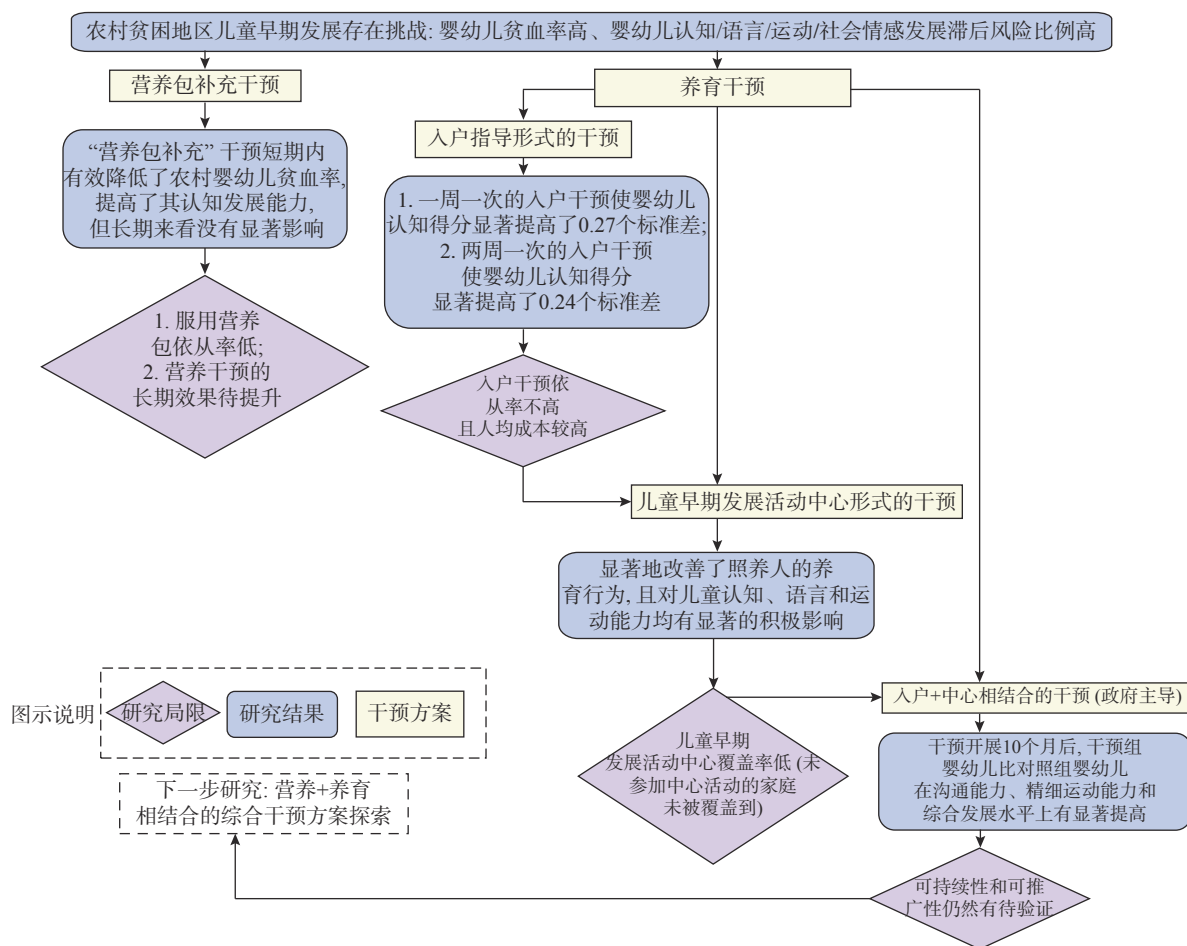


图5 “儿童早期发展”领域 RCT 研究过程图

### (一) 随机干预实验在“教师与教育”领域的应用

#### 1. 研究问题

近年来,我国在教育领域取得了长足发展,但城乡之间在办学水平和教学质量等方面仍然存在差距。从教育的结果指标来看,城乡间教育水平的不均衡至少体现在以下三个方面:首先,城乡居民总体受教育水平差距较大,2010年人口普查数据显示,农村人口平均受教育水平为7.58年,比城市人口低了3年(孙炜红,张冲,2014)。其次,农村学生高中录取率和大学录取率显著低于城市学生,重点学校录取率的差距更大(Liu et al., 2009; 杨倩,谢作翔,2017)。再次,学生的学业表现是衡量教学质量的一个重要指标(Glewwe et al., 2006),在这一指标上农村学校学生显著落后于城市学校学生(Liu et al., 2009; 王云峰等,2012; 中国教育报,2016)。

为补齐农村教育的短板,促进城乡教育优质均衡发展,国家实施了一系列措施来改善农村学校办学条件、提高农村学生在各个阶段的入学率,如2001年起实施的“农村小学撤点并校”和2006年起实施的“免除农村义务教育学杂费”等政策。教育是一个复杂的系统工程,不同时期农村教育面临的主要问题有很大差异,导致农村教育相对薄弱的原因也很多。过去十多年间大量关于中国农村中小学阶段教育的随机干预实验研究重点关注了如下几个问题:

(1)寄宿制学校问题。自2001年开始,我国对农村小学进行了布局结构调整,即将村小合并到规模较大的镇中心小学。从2001年到2005年,每年合并了超过2万所农村小学(国家统计局,2001; 国家统计局,2010)。与此同时,也产生了大量的寄宿生。农村小学中寄宿、半寄宿制小学所占比重逐年增加,很多学校寄宿生的数量超过50%(东方网,2005)。国内外研究发现,与非寄宿生相比,寄宿生具有更多的不良行为和更差的学业表现(Malcolm, 1970; 庞丽娟,2005; Moswela, 2006; 熊向明,2007; Adams,



2007; 叶敬忠, 2008; Luo et al., 2009; Mo et al., 2013)。与此相关的是农村小学寄宿生的管理问题。2007 年对 144 所农村小学的一项调研发现, 仅有 5% 的寄宿制学校有专职的宿舍管理员, 其他都是由教课老师兼任生活老师, 而这些生活老师几乎没有接受过任何正式的宿舍管理方面的培训(Luo et al., 2009)。

(2)非义务教育阶段的学费问题。早在 2006 年, 国家就实施了免除农村义务教育学杂费等政策; 到 2008 年, 又进一步在全国范围内全部免除了义务教育阶段的学杂费。然而, 在非义务教育阶段, 上学的经济成本仍然是阻碍贫困学生获得受教育机会的重要原因。1997 年到 2006 年, 高中学费从平均 1620 元上涨到平均 4500 元(Liu et al., 2011)。通过分析 41 个国家高中学费的国际比较数据和针对我国西北某省农村地区 1100 多名学生的调查数据, Liu 等人(2009)认为高额学费可能会阻碍农村学生入读高中。此外, 由于较高的升学压力, 对贫困学生而言, 高额学费带来的低升学预期还可能降低他们准备升学考试的积极性, 从而影响其学业表现, 并进一步降低他们的竞争力。

(3)师资问题。一直以来, 我国农村学生的学业表现远落后于城市学生, 城乡学生的学业表现存在较大差距(王云峰等, 2012; 中国教育报, 2016)。学生学业表现受到多种因素影响, 其中教师是核心因素之一(Sanders et al., 1997; Rockoff, 2004; Nye et al., 2004)。为了提高教师的工作积极性、体现“多劳多得, 优绩优酬”的激励理念, 我国出台了《关于义务教育学校实施绩效工资指导意见》, 要求从 2009 年起在全国义务教育学校实施教师绩效工资, 并规定绩效工资包括 70% 的基础性工资(固定部分)和 30% 的奖励性绩效工资, 由学校按照教师工作量和成果贡献制定具体的绩效考核标准(教育部, 2008; 国务院, 2009; 王聪, 2017)。研究者发现绩效工资政策的实施增加了教师的工资福利(付卫东等, 2010; 王聪, 2017)。但是绩效工资具体实施方案五花八门, 效果也参差不齐。尤其是部分农村学校的绩效工资方案, 并没有真正体现“多劳多得, 优绩优酬”, 更没有起到激励教师的作用(范先佐等, 2011; 付卫东等, 2011; 叶怀凡, 2016; 胡耀宗等, 2017)。

围绕中小学阶段农村教育质量和学生人力资本质量的提升问题, 学者们结合国际经验和中国农村的实际情况, 设计和实施了多项随机干预实验。这些随机干预实验有针对教师的干预, 也有针对学生的干预, 项目的抽样方式及调研形式也略有不同, 但均在农村贫困地区实施以保证研究对贫困农村人力资本培育的针对性。具体来说, 可以分为“寄宿制学校生活老师培训”“贫困生事先资助承诺”和“教师绩效激励”三大类, 表 1 是对 7 个实验研究的时间、样本量、干预方式和干预结果的简要介绍。

表 1 农村中小学“教师与教育”领域随机干预实验研究结果汇总表

序号	时间	地点	干预方式及样本量	干预效果	文章来源
1	200709-200906	陕西	样本: 共10个县10所中学, 768名寄宿生 对照组: 5所学校, 294名寄宿生 干预组: 5所学校, 39名生活老师, 474名寄宿生 干预内容: 培训管理寄宿生的生活老师	生活老师培训显著减少了寄宿生的不良行为。其中学生迟到或者早退行为减少了48%、课后不良行为(吵架、打架)减少了78%	Yue et al., 2014
2	200703-200706	陕西	样本: 共8个县10所高中的592名高三贫困生 对照组: 326名贫困生 干预组1: 66名贫困生 干预内容: 高考前三个月得到5000元事先资助承诺 干预组2: 67名贫困生 干预内容: 高考前三个月得到2500元事先资助承诺 干预组3: 66名贫困生 干预内容: 高考后一天得到5000元事先资助承诺 干预组4: 67名贫困生 干预内容: 高考后一天得到2500元事先资助承诺	相对于对照组: 所有干预组的高考成绩没有不同; 所有干预组的录取比例没有不同; 干预组1报考师范类院校和国防类院校的比例显著降低	Liu et al., 2011
3	201011-201109	陕西河北	样本: 共15个县30所小学, 380名九年级贫困生 干预内容: 每班4名贫困生配为两对, 随机给予每对中的一名学生1500元事先资助承诺, 共计190名学生得到干预	一年后: 贫困生总体高中入学率增加了7.9个百分点, 职业高中和普通高中入学率没有显著改变	Yi et al., 2015

续表 1

序号	时间	地点	干预方式及样本量	干预效果	文章来源
4	201011-201305	陕西河北	样本: 共15个县132所学校, 1892名七年级贫困生 纯对照组: 66所学校 干预组: 66所学校 干预内容: 将干预组学校每班4名贫困生配为两对, 随机给予每对中的一名学生事先资助承诺, 共计474名贫困生得到干预(称为“配对对照组”), 其他474名贫困生未得到干预	一年后: 贫困生计划就读普通高中的比例增加了15.2个百分点, 就读职业高中的比例没有增加, 学习成绩没有改善 三年后: 贫困生的高中入学率(普通高中和职业学校)没有改善; “配对对照组”贫困生和纯对照组贫困生在上述指标上均没有显著区别	Yi et.al, 2015; Li et al., 2017
5	201309-201405	陕西甘肃	样本: 共16个县216所小学, 243名教师, 7373名学生 对照组: 57名教师, 1863名学生 “绝对值”干预组: 62名教师, 1738名学生 “增加值”干预组: 62名教师, 1996名学生 “增加值百分位”干预组: 62名教师, 1776名学生	“增加值百分位”绩效激励, 能够显著提高学生的学业表现0.15个标准差, 尤其是将学困生的学业表现提高0.19个标准差。“增加值百分位”绩效激励方式更能有效促进教师改变教学行为, 比如讲授更多中等偏难的内容	Loyalka et al., 2019
6	201409-201505	陕西甘肃	样本: 共16个县103所小学113名教师3411名学生 对照组: 57名教师, 1640名学生 “增加值百分位”干预组: 56名教师, 1771名学生	“增加值百分位”绩效激励, 能够显著提高学生的学业表现0.10个标准差, 将学困生的学业表现提高0.15个标准差。“增加值百分位”绩效激励方式促使教师在教学内容上同时考虑学困生和优等生的需求。教师对学生的关心和与学生交流意愿增强	Chang et al., 2019
7	201510-201606	陕西甘肃	样本: 共16个县200所中学276名教师12095名学生 对照组: 131名教师, 5785名学生 “增加值百分位”干预组: 145名教师, 6310名学生	对学生的学业表现没有显著改变, 也未对教师的教学行为产生影响	结果仍在分析中, 文章尚未发表

2. 农村中小学“教师与教育”领域的随机干预实验研究

研究一: 农村寄宿制学校生活老师培训干预

研究简介

为探索解决寄宿制学生问题的有效方案, 研究者于 2007 年到 2009 年在陕西省进行了关于寄宿制学校生活老师的随机干预实验。从 144 所学校中随机抽取 10 所学校(学生样本共 2294 名, 其中寄宿生 768 名), 随机分配 5 所为干预组(474 名学生), 另外 5 所为对照组(294 名学生)。

研究团队联合教育学、心理学和营养健康方面的专家共同开发了适合农村生活老师的培训教材、练习手册和教学计划, 培训内容共分为 10 个模块(学生来校登记、晨起锻炼及洗漱、就餐、上午上课期间、午休、下午上课期间、下午课外活动、晚饭后活动、晚上睡觉、健康)。通过讲解、游戏、互动等方式对 39 名干预组生活老师进行培训。最终, 通过学生的一系列行为变化来衡量干预效果, 主要包括三个方面: 不良行为(学生上课说话、不注意听讲); 缺课(逃课、迟到、早退、吵架、打架); 健康问题(学生是否感冒发烧、是否拉肚子等)。

研究结果

研究发现, 生活老师培训显著减少了寄宿生的不良行为, 其中迟到或者早退行为减少了 48%, 课后不良行为(吵架、打架)减少了 78%, 但是对寄宿生上课期间的不良行为和学业表现没有显著影响(Yue et al., 2014)。

在该研究进行的过程当中,研究团队发现,寄宿制学生面临的问题可能只是农村学校布局结构调整的一个表现。事实上,在开展该研究的过程当中,研究团队也观察到了一些农村小学亟待解决的其他问题,比如学生营养健康问题、心理健康问题和学业表现落后问题等,这些问题也是研究团队后续开展一系列营养健康与教育相关研究的基础。

### 研究局限

寄宿制学校生活老师培训项目是研究团队开展的第一个中国农村教育领域的 RCT 项目,由于是第一次探索,研究设计存在一定的局限性,主要表现在该项目仅在 10 所学校进行,干预组只有 5 所农村学校,可能存在统计功效不足的问题,从而影响了该研究的外部有效性。

### 研究二:针对高中生的事先资助承诺干预

#### 研究简介

针对农村贫困学生可能因为经济原因选择不上大学的问题,研究团队于 2007 年—2008 年在陕西省 8 个贫困县的高中三年级开展了面向贫困学生的事先资助承诺实验。2007 年,研究团队对 1177 名高二学生进行了基本情况调查,获取了样本学生的家庭主要资产清单,并根据清单中物品的时价估算了每个学生的家庭资产,然后以低于 7600 元为标准选出农村贫困生<sup>⑧</sup>,样本学校中的 592 名学生被确认为贫困生。

2008 年春,研究团队对所有贫困生样本进行了随机分配,进行了所有干预组和对照组样本的基线控制变量的平衡性检验。约一半贫困生被分入干预组,研究团队向他们进行了事先资助承诺干预,即:如果学生被本科一批/二批高校录取即可得到资助。干预方式按照事先资助承诺金额和时间的不同分为四种:金额上,一半人得到 2500 元的资助承诺,一半人得到 5000 元的资助承诺;时间上,一半人在 3 月初(春季学期开学第一天)获得承诺,一半人在 6 月(高考结束后的第二天)获得。

#### 研究结果

研究团队对学生的高考成绩进行了回归分析,发现相比没有获得资助承诺及 6 月才获得资助承诺的贫困生,3 月即获得资助承诺的贫困生没有获得更好的高考成绩(Liu et al., 2011)。同时,由于所有被本科一批/二批录取的样本学生都决定上大学,因此资助也没有改变学生是否上大学的决定。事先资助承诺干预的影响主要体现在志愿填报上:3 月份资助组学生报考有学费减免和补助的师范类院校和国防类院校的比例显著降低,也就是说事先资助承诺使得原本的激励机制发生了变化。由于离填报志愿的截止日期太短,6 月份资助组学生填报的志愿没有明显不同。

#### 研究局限

事先资助承诺干预对学生的高考分数没有影响,可能是由于资助承诺给得太晚,资助产生的激励效果在三个月时间内无法显著影响分数;也可能是由于上大学的边际收益很高,导致资助对努力程度的边际作用太小。因此,在后续的实验中需要改进研究设计。此外,能够坚持到高三下半学期的学生可能已经做好了一定要上大学的心理准备,因此资助对最终是否上大学没有影响。另外,也可能是研究分析中没有考虑溢出效应,比如,很有可能干预组学生获得的事先资助承诺对对照组学生也产生了影响。

### 研究三:针对初中生的事先资助承诺干预

#### 研究简介

2010 年,研究团队探索了事先资助承诺干预对陕西省和河北省的 15 个国家级贫困县农村九年级贫困学生升入高中的影响。吸取了此前研究团队开展的高中三年级学生事先资助承诺干预研究的经验,这次干预研究选择了在秋季学期实施。通过基线调查、随机分配和平衡性检验等过程,研究团队选出了干预组每个班级中最贫困的 4 名学生,并根据 4 名贫困生的基础特征配为两对,随机给予每对中的一名学生事先资助承诺。2010 年 12 月,项目组召集了所有九年级干预组学生及其家长/监护人并与其签订了事先资助承诺合同。合同规定,如果该学生在 2011 年 9 月之前考入普通高中或职业学校,将得到为期三年、每年 1500 元的奖励。



基于研究团队之前开展的高中贫困生事先资助承诺干预项目的经验,即干预更早开始会获得更好的效果,在开展这项针对九年级贫困生的研究项目时,研究团队也在同一地区扩充了样本,增加了针对七年级学生的事先资助承诺干预,在设计方案时也更好地考虑到了溢出效应。在研究项目具体执行阶段,研究团队将132所学校平均分为两组:第一组为“事先资助承诺学校”,这些学校中的学生有机会获得事先资助承诺;第二组为对照组学校,这些学校中没有学生会获得事先资助承诺。然后,进一步将“事先资助承诺学校”中的贫困生根据是否获得事先资助承诺随机分为两组。每班各有2名学生进入配对干预组,即获得事先资助承诺;另外2名进入配对对照组,即不会获得事先资助承诺(根据基线调查确保两组学生的平衡性)。2010年11月,研究团队与配对干预组的学生签订了事先资助承诺合同,合同规定,如果该学生在2013年9月之前考入职业学校或普通高中,研究团队将每年给该生提供1500元的奖励。之后,研究人员每年大约在5月左右致电每位学生以提醒他们承诺一直有效。

### 研究结果

针对九年级组样本学生的干预结果,研究团队分析发现,事先资助承诺干预将贫困生的总体高中录取率(包括普通高中和职业高中)提高了7.9个百分点,但仅在10%的水平上显著;分别看普通高中和职业高中的录取率则发现干预没有显著影响(Yi et al., 2015)。研究人员分析,可能的原因与高中生事先资助承诺研究项目相同:能够坚持到初三的学生可能已经做好了一定要上高中的心理准备,因此资助对能否上高中影响很小。

而对七年级组样本学生的干预结果进行分析,可以发现,事先资助承诺对贫困生一年后的成绩没有任何影响;而且,虽然该项目显著增加了贫困学生在一年后对于上高中的期待,但是对三年后真实的高中入学率(包括普通高中和职业学校)没有显著影响(Yi et al., 2015; Li et al., 2017)。结合针对九年级学生的研究结果,似乎可以得到无论事先资助承诺何时给出对学生产生的影响都极其微弱这一结论。

### 研究小结

总体而言,针对不同群体、干预时长不同、干预强度不同、考虑到溢出效应的一系列事先资助承诺实验得出较为一致的结论:在中国农村,事先资助承诺既不能改善学生的学习成绩,也没有显著提高他们的升学率。考虑到事先资助承诺在哥伦比亚、巴基斯坦、墨西哥和巴西等发展中国家取得了一定成效,提高了学生的入学率,事先资助承诺在中国农村地区为何会失败尤其值得研究人员关注(Schultz, 2004; Heinrich, 2007; Nazmul & Parajuli 2008; Barrera-Ororio et al. 2011)。

研究团队分析,一个可能的解释是,干预提供的资助金额还不够高,不足以激发行为改变。中国农村地区的学生面临相当大的机会成本,2012年典型非技术工人的收入约为每月2900元(国家统计局, 2014),每年仅有1500元的资助可能不足以产生明显的效果。然而,国际经验表明,即使资助的金额不高,项目往往也会有效。在初中项目中使用的资助金额(每年190美元)大约等于(甚至高于)已证明在其他发展中国家有效的事先资助承诺项目,例如柬埔寨的每年45美元至60美元和墨西哥的每年200美元至250美元;并且,尽管不同项目的资助金额差异很大,都还是产生了实质性影响(De Janvry & Sadoulet, 2004; Filmer & Schady, 2009)。

另一个解释是,农村贫困学生在继续上学方面面临许多挑战,只提供资金方面的帮助无法解决问题。许多研究发现,提供以学生入学为条件的事先资助承诺并不能转化为更高的学业成绩(Behrman et al., 2005; Banerjee et al., 2007)。由于贫困常与学业表现相关,即使事先资助承诺提供了经济支持,也可能无法帮助学生提高成绩,从而无法帮助学生升入更高一级的学校。在我国,由于普通高中的学生竞争激烈,某些学生可能会由于感知到升学机会太低而放弃学习;相对地,虽然职业高中录取要求低,但研究表明,职业高中通常教学质量较低,不为学生所信任(Loyalka et al., 2013)。研究人员分析,如果这些问题在学生决策中起着重要作用,那么仅提供金钱就无法从根本上解决问题。

研究团队总结了这一系列干预的经验后认为,在我国,单纯提供资助的方式极有可能达不到预期效果,我们可能更需要从教育体系的供给侧入手,通过提高学生学业成绩来大幅度提高学生升入高中

的预期录取率和实际录取率。研究团队也积极与地方政府合作实施了一些政府主导的教育研究项目。例如,陕西省N县于2009年实施了高中免学费政策,承诺前500名被录取的初中毕业生可以免除三年高中学费(每年1500元人民币)。研究团队将与N县基础特征较为相似的邻县作为对照组,研究发现,经过一年的学习,N县的学生成绩平均比对照组多提高了3.1分,约等于0.2个标准差。通过异质性分析发现,项目的主要影响发生在家庭资产最低的20%学生身上,与其他学生相比,他们的成绩多提高了3.8分(Chen et al., 2013)。此外,该项目也将N县初中毕业生升入普通高中的比例提高了21个百分点,进入职业高中的比例降低了7个百分点,直接进入劳动力市场的比例降低了11.9个百分点,第二年再次参加中考的可能性降低了2.1个百分点(Bai et al., 2017)。

#### 研究四：“优绩优酬”——农村教师绩效激励实验研究

##### (1)教师绩效激励第一期

##### 研究简介

研究团队于2013年到2014年在西北2省进行了关于农村教师绩效激励项目的实验,共收集了216所农村完全小学的243名教师和7373名学生的信息,干预对象为六年级数学教师。研究人员将216所学校随机分成4组,其中52所学校为对照组(57名教师,1863名学生),54所学校为“绝对值”干预组(62名教师,1738名学生),56所学校为“增加值”干预组(62名教师,1996名学生),54所学校为“增加值百分位”干预组(62名教师,1776名学生)。同样地,分组完成后也进行了平衡性检验,保证干预组和对照组在统计上没有显著差异,从而形成统计意义上的“双胞胎”。

“绝对值”干预组是根据评估调查时学生的标准化数学测试成绩来计算每位样本教师所教学生的平均数学成绩;“增加值”干预组则是根据学生评估调查的标准化考试成绩与基线调查的标准化考试成绩之差,得到每位学生成绩的增加值,再计算每位样本教师所教全部学生成绩增加值的平均数;“增加值百分位”干预组是先计算学生评估调查和基线调查的标准化数学测试之差,再从该组全体样本中找出基线调查成绩一样的学生(即“起点相同的学生”),根据这些学生考试成绩的增加值对起点相同的同学进行百分位排名,每个样本学生都获得一个百分位排名,最后再对每位样本教师所教全部学生的增加值百分位取平均值,将其作为该教师的教学质量衡量指标。干预组样本教师的激励奖金取决于其教学质量的百分位,三组干预教师分别按照教学质量进行百分位排名。在第一期教师绩效激励项目中,百分位最高的教师能够从项目组得到7000元的激励奖金,百分位每降低一名,激励奖金就减少70元。

根据上述设计方案,研究团队开发了干预协议和协议指南,包括项目背景、激励教师的方式、激励奖金的计算办法、双方的权利和义务等。研究人员根据每位样本教师被分配到的不同组别,分别对干预组教师进行了现场集中培训,向他们介绍项目背景、协议指南并签订激励协议。最终,研究团队根据基线调查和评估调查的学生学业表现结果,计算出每一名教师的教学质量百分位排名和其应得的激励奖金数额,并向样本教师发放激励奖金。

##### 研究结果

研究团队对干预结果进行分析后发现:无论哪种激励方式,都有助于提高学生的学业表现;但是只有“增加值百分位”的绩效激励方式,才显著将学生的学业表现提高0.15个标准差,尤其是将学困生的学业表现提高了0.19个标准差。并且,研究还发现,“增加值百分位”的绩效激励方式更能有效促进教师改善教学行为,比如在授课过程中既讲授中等难度的内容也包含更难的内容,既照顾了在学习上有困难的学生,也考虑了优等生的需求(Loyalka et al., 2019)。

##### 研究局限

该研究的结论是只在进行了一年干预后得出的,研究人员也在思考基于成绩增加值百分位的激励方案对学生成绩的促进作用是偶然事件还是长期效果。从国际上看,在肯尼亚、印度和以色列进行的研究表明,教师激励对学生的学业表现有长期影响(Muralidharan, 2011)。那么在我国西北地区,如果

开展长期干预,基于成绩增加值百分位的激励方案是否依然有助于提高学生学业表现?这也是值得研究者探究的问题。

## (2)教师绩效激励第二期

### 研究简介

为了验证教师绩效激励的长期影响,研究团队在上述研究项目的基础上,于2014年到2015年进行了第二期教师绩效激励实验。本期实验研究仍然沿用第一期研究的样本学校,其中“增加值百分位”干预组共57名教师1640名学生,对照组共56名教师1771名学生,实验的过程与教师绩效激励干预第一期一致。

### 研究结果

研究人员发现,以“增加值百分位”的绩效激励方式对农村教师进行干预,干预一年能够显著将学生的学业表现提高0.10个标准差,将学困生的学业表现提高0.15个标准差。另外,对教师教学行为进行了测量后发现,“增加值百分位”的绩效激励方式更能有效促进教师改变教学行为。除了在教学内容上同时考虑学困生和优等生的需求,更为重要的是,教师对学生的关心和与学生的交流增多,并且切实采取了一些措施来改善教学行为和方式。这可能是“增加值百分位”绩效激励方案能够提高学生学业表现的内在机制(Chang et al., 2020)。

## (3)教师绩效激励第三期

### 研究简介

研究团队通过上述两期研究已经得出了教师绩效激励在小学阶段长期有效的结论。再长远一点看,这样的激励方式在初中阶段是否同样有效?为此,研究人员继续开展了教师绩效激励第三期研究。实验于2015年至2016年在西北2省3市随机选取16个县200所农村初中学校进行,总样本共276名七年级数学教师,其中干预组采取“增加值百分位”的激励干预方式(145名教师样本),对照组不进行任何干预(131名教师样本),激励方案和实施步骤与教师绩效激励研究第一期和第二期保持一致。

### 研究结果

研究人员发现,“增加值百分位”的教师绩效激励干预方式对农村初中学生的学业表现没有产生显著影响,也未对教师的教学行为产生影响。教师绩效激励干预在小学阶段有效,而在初中阶段无效,其可能的原因和影响因素仍在分析当中。

## (4)小结

三期教师绩效激励的研究结果表明,在农村小学阶段对教师进行“增加值百分位”绩效激励,能够显著提高学生的学业表现0.10—0.15个标准差,尤其是将学困生的学业表现提高了0.15—0.19个标准差。在分析其作用机理后,研究人员发现“增加值百分位”的绩效激励方式通过显著改变教师的教学行为,进而改变了学生的学业表现。

在现有的农村教师绩效工资体系下,至少在小学阶段,教师的绩效不仅要与学生的学业表现挂钩,而且要与学生的增加值百分位挂钩。也就是说,教师如果想获得更高的绩效工资,对班级内不论是那些预期成绩进步空间更大的学困生和后进生,还是中等生和优等生,都要予以关注,从而真正体现有教无类、多劳多得、优绩优酬。在当前我国农村学生相对于城市学生受教育水平较低、机会较少的大背景下,这个方案最能引导教师关注发展起点各不相同的农村孩子,给所有孩子公平发展的均等机会。

## 3. 小结

近年来,我国不断提高财政性教育经费支出占国民生产总值的比例,不断提高对农村教育的投资,努力解决城乡间的教育不均衡问题,以实现教育均衡发展。然而,限于种种现实原因,农村学生的受教育机会和受教育质量的现状仍不容乐观。政策制定者在增加教育投资的基础上,需要优化对农村教育的投资结构,制定符合农村实际的教育政策。在各项学校投入中,教师是影响学生受教育质量的重要因素,对农村教师进行培训,设计和推广行之有效的教师激励方案,可能是解决农村教育问题的关键。



## (二) 随机干预实验在“信息技术与教育”领域的应用

### 1. 研究问题

自2012年国家提出深入推进义务教育均衡发展基本目标起,在办学经费得到保障的前提下,不论城乡我国每一所学校均建设成为了符合国家办学标准的学校。除教育基础设施外,教育资源也基本可以满足学校教学的需要,包括开齐国家规定课程、教师配置更加合理及教师整体素质极大提高等。以往诸如教学设施不完善、师资缺乏、教学质量较低等制约城乡教育均衡发展的的问题正在逐步得到解决(国务院,2012)。

在实现教育均衡发展目标的进程中,一个不容忽视的挑战是农村学生仍缺乏有效的课后补习、辅导等教育资源。城市学生学业成绩落后时,可以获得教师、家长的及时帮助,或寻求商业补习机构的付费补习辅导服务。然而,农村学生往往很少有这样的资源。农村教师不仅需要完成日常繁重的教学任务,还需要承担部分行政管理工作,很少有机会为学生开展课外补习辅导。此外,由于农村学生父母大多受教育程度低或外出打工、农村地区商业补习辅导昂贵且稀缺,城市学生常见的课后补习(教师辅导、家长辅导和商业辅导)对农村学生来说几乎都是欠缺的(曾满超等,2010; Lai et al, 2015a)。

从国际经验来看,计算机辅助学习(Computer Assisted Learning, CAL)具有简便易行、易于推广的优点,或许可以成为农村学生课后补习的一种替代性选择(Banerjee et al., 2007; Linden, 2008; Barrow et al., 2009)。从广义上来讲,与教育活动相关的在计算机上实现教与学的软件项目都可以称为“计算机辅助学习项目”(Escueta et al., 2017)。计算机辅助学习不同于以往发放计算机设备的项目,它不涉及提供硬件设备,而是着眼于在已有计算机设备上使用开发好的特定软件包(Rouse & Krueger, 2004)。使用计算机辅助学习软件与在线课程不同的地方在于,计算机辅助学习旨在开发有特定目的的软件包以提升某些技能(如提高数学能力、阅读理解能力等),是一种“智慧补习系统”,可以帮助学生弥补在知识点和技能上的缺陷(Escueta et al., 2017)。

国内外已有诸多关于计算机辅助学习对学生学业表现影响的研究。发达国家在过去十多年中涌现了一批关于计算机辅助学习影响效果的研究,比如,美国的很多研究发现计算机辅助学习在学校教学中的应用对学生学业成绩产生了有利影响(Fuchs & Woosmann, 2004; Goolsbee & Guryan, 2006)。Barrow等(2009)发现计算机辅助学习在芝加哥的实验项目中使学生的数学成绩提高了0.17个标准差。Banerjee等人(2007)的研究表明,计算机辅助学习可以显著提高学生的数学成绩,即使在项目结束一年后再看项目效果,学生成绩仍有明显增长。Sun等人(2008)和Hwang等人(2012)的研究表明,计算机辅助学习与传统课堂学习相比,除能使学生成绩明显提高外,75%的学生表示出对计算机辅助学习的兴趣和喜爱。He等人(2008)和Mo等人(2014a)的研究表明在发展中国家开展计算机辅助学习项目对学生成绩有着积极影响。

但也有研究发现使用计算机辅助学习对学生成绩产生了消极影响或没有显著影响。Angrist和Lavy(2002)的研究发现在以色列将计算机教学融入日常教学中导致了八年级学生数学成绩的下降。Dynarski等人(2007)以及Rouse和Krueger(2004)发现在美国开展的计算机辅助学习项目没有显著提高学生的数学成绩和阅读能力。Ritchie和Newby(1989)发现计算机辅助学习并没有明显提高学生学习成绩,但使学生学习态度发生了积极改变。Richardson和Long(2003)的研究表明,如果没有课程老师帮助学生使用计算机辅助学习软件,学生成绩不会明显提升。

综上可知,世界上包括发达国家和发展中国家在内的不同国家已先后开展了不同形式的计算机辅助学习干预项目,研究结果也不太一致。但大部分研究发现计算机辅助学习有利于提升学生学业表现,且能够为学生带来其他方面的有益影响(比如提高学习兴趣等)。国内外的这些研究为在中国开展有关信息技术与教育的干预实验提供了丰富的实践经验和启示,在借鉴已有研究的基础上,国内研究者和教育政策制定者也需要结合本土实际情况,探索适合中国发展现状的信息技术与教育干预模式,以有效改善现代信息技术在农村教育中的应用状况。

《国家中长期教育改革和发展规划纲要(2010—2020)》明确提出:“到2020年,基本建成覆盖城乡各级各类学校的教育信息化体系,促进教育内容、教学手段和方法现代化。”(教育部,2010)《教育信息化“十三五”规划》指出:“各级教育行政部门要保障基础性数字教育资源的供给,并发挥好已有资源的作用,利用以互联网为主的多种手段将资源提供给各类教育机构,尤其是农村、边远、贫困、民族地区的学校免费使用。”(教育部,2016)教育部《教育信息化2.0行动计划》指出:“各地将教育信息化作为重要指标,纳入本地区教育现代化指标体系。全面开展面向区域教育信息化的督导评估和第三方评测,提升各地区和各级各类学校发展教育信息化的效率、效果和效益。”(教育部,2018a)

在这样的背景下,计算机辅助学习能否提高中国农村贫困地区学生的学业成绩?是否能成为教育扶贫的精准工具?如果有效,其影响机制是什么?是否具备易操作、可复制的推广价值?这一系列问题的解决尚缺乏可靠的实证依据。为了尝试回答这些问题,中国从事农村教育的研究团队自2011年至2016年在全国13个省(市)的农村学生中先后开展了7项大型随机干预实验研究,努力从学生学业表现和非学业表现等角度识别计算机辅助学习对学生学业表现的影响及其具体机制,并探讨计算机辅助学习在中国农村是否具有大规模推行的科学价值和现实意义。

2. “信息技术与教育”领域的随机干预实验研究

通过对这7项随机干预实验(具体信息见表2)进行梳理,不难发现其中的诸多相似之处。首先,从样本选取对象来看,样本学校主要是农村学校或城市打工子弟学校,均属于学业表现相对较弱的学生群体。其次,样本选取流程基本一致:通过计算统计功效确定需要选取县或区的数目,再从省份中随机抽取县区并联系地方教育部门获取当地农村学校基础信息,最后从所有农村小学中选取三年级以上学生纳入样本框(考虑到国家规定的信息技术课只对三年级以上学生开设,为确保样本学生对计算机操作有一定的掌握程度,只选择三年级以上学生)。最后,7项实验都严格按照随机干预实验的步骤进行后续实验。

表2 “计算机辅助学习”随机干预实验汇总表

序号	时间	地点	干预方式与样本量	实验效果	来源
1	2010.12-2011.10	北京打工子弟学校	样本:794名3年级学生中抽取300名学生 对照组:25个班,每班6人,共150名学生 干预组:25个班,每班6人,共150名学生 干预内容:给干预组学生分发装有单机版数学科目的计算机辅助学习软件的笔记本电脑,同时对学生和家长开展培训 干预时长:6个月	相对于对照组,干预组学生的电脑技能提高0.33个标准差;数学成绩提高0.17个标准差;使用电脑的时间增加,相应减少了看电视时间;学生自我效能感也有所提升	Mo et al. (2013)
2	2010.9-2011.12	北京打工子弟学校	样本:43所北京打工子弟学校中抽取24所学校,3年级,共2514名学生 对照组:30个班级,1447名学生 干预组:24个班级,1067名学生 干预内容:每周2次40分钟数学单机版CAL课程,时间安排在中午或下午放学后 干预时长:1个学年	相对于对照组,CAL干预提升了学生数学成绩0.15个标准差,并显著增加了学生的学习兴趣;且父母受教育程度低的学生受益更明显	Lai et al. (2015a)
3	2011.3-2011.6	陕西农村寄宿制学校	样本:72所农村学校中2726名3年级和5年级寄宿学生 对照组:36所学校1451名学生(602名3年级+849名5年级) 干预组:36所学校1275名学生(553名3年级+722名5年级) 干预内容:每周2节40分钟数学单机版CAL课程,时间安排在校内课外时间 干预时长:1个学期	相对于对照组,CAL干预提升了学生数学成绩0.12个标准差;且对贫困家庭学生效果更显著	Lai et al. (2013)

续表 2

序号	时间	地点	干预方式与样本量	实验效果	来源
4	2011.3- 2011.6	青海少数民族地区学校	<b>样本:</b> 57所学校1889名3年级学生 <b>对照组:</b> 31所学校, 3年级1125名学生 <b>干预组:</b> 26所学校, 3年级764名学生 <b>干预内容:</b> 每周2节40分钟语文(汉语)单机版CAL课程, 时间安排在校内课外时间 <b>干预时长:</b> 1个学期	相对于对照组, CAL干预提升了学生语文成绩0.14—0.20个标准差; 提高数学成绩0.22个标准差; 且对学生自我效能感有正向影响	Lai et al. (2015b)
5	2011.9- 2012.6	陕西农村学校	<b>样本:</b> 72所农村学校, 5267名3年级和5年级学生 <b>对照组:</b> 36所学校, 3年级学生1212名, 5年级学生1620名 <b>干预组:</b> 36所学校, 3年级学生1067名学生, 5年级学生1368名 <b>干预内容:</b> 每周2节40分钟数学单机版CAL课程, 时间安排在校内课中(占用信息技术课时间) <b>干预时长:</b> 1个学年	相对于对照组, CAL干预提升了3年级学生数学成绩0.17个标准差; 5年级学生数学成绩提升0.17个标准差	Mo et al. (2014a)
6	2013.9- 2014.6	青海农村学校	<b>样本:</b> 127所农村学校, 5年级6304名学生 <b>对照组:</b> 4000名学生 <b>干预组1:</b> 计算机辅助教学(Computer Aided Instruction, 简称CAI)干预组, 22所学校, 1236名学生 <b>干预组2:</b> CAL干预组, 22所学校, 1068名学生 <b>干预内容:</b> CAI组提供不与课程结合的单机计算机辅助学习, CAL组提供与课程结合的单机计算机辅助学习, 干预科目均为英语。CAL安排在校内课中时间 <b>干预时长:</b> 1个学年	相对于对照组, CAI没有显著提升学生的英语成绩, 但是CAL显著提升了学生的英语成绩0.08个标准差	Bai et al. (2016)
7	2015.12- 2016.6	陕西农村小学为主, 兼顾12个省份样本	<b>样本:</b> 44所农村学校, 5年级1650名学生 <b>对照组:</b> 22所学校, 936名学生 <b>干预组:</b> 22所学校, 714名学生 <b>干预内容:</b> 干预组接受在线英语科目计算机辅助学习, 每周2次40分钟英语OCAL课程, 时间安排在校内课中 <b>干预时长:</b> 1个学期	相对于对照组, CAL干预显著提升学生英语成绩0.56个标准差; 同时学生对英语学习的态度变得积极, 教育期望提高	Bai et al. (2018)

### 研究一: “计算机辅助学习”干预在打工子弟学校学生中的有效性验证

#### 研究简介

国际经验表明, 给学生分发电脑不会提高学生的学业成绩(Goolsbee & Guryan, 2006; Fairlie & Robinson, 2013), 增加网络接入、提升宽带速度也不会提高学生学业成绩(Faber et al., 2015), 电脑的分发、使用应当与学校的课堂教学结合起来(Escueta et al., 2017)。因此, 在中国开展的多项信息技术与教育领域的随机干预实验就是希望在借鉴国际经验的基础上, 将课程融入计算机软件中, 以更好地开展计算机辅助学习的干预效果研究。

表2中的实验1是中国第一个“信息技术与教育”领域的随机干预实验研究项目(Mo et al., 2013b), 该项目旨在缩小信息技术鸿沟, 改善农村学生的学业表现。2010年至2011年, 研究团队将样本中50个班级的300名北京打工子弟学校学生分为2个组, 给干预组(25个班)学生提供免费的笔记本电脑(一人一台, 笔记本电脑中已事先安装了计算机辅助学习软件), 并培训学生和家长使用计算机辅助学习软件的方法, 对照组(25个班)学生则不做任何干预。

研究团队也在同一时间开展了实验2, 该实验旨在缩小随父母来到北京就读打工子弟学校的农村学生与就读高质量城市公立学校的城市学生之间存在的教育差距。许多流动学生(父母由于务工经常



搬家,他们在许多不同的学校进进出出)面临的重大问题之一是他们经常学业落后,并且很难追赶上。由于许多流动家庭学生在学校学业落后,实验2的主要目标是为他们提供课后辅导。研究团队将24所学校2514名学生随机分为干预组和对照组,在与正常的数学课程不冲突的时间内(例如,午休时、下午放学后),组织干预组学生在学校机房开展每周2次、每次40分钟的数学单机版CAL课程,以评估计算机辅助学习是否对学业表现落后的学生有正向影响,是否有助于缩小他们与公立学校普通城市儿童的成绩差距(Lai et al., 2015a)。

### 研究结果

基于实验1的研究结果,研究团队发现发放笔记本电脑提升了学生使用电脑的时间,同时减少了看电视的时间,学生的基本电脑操作技能显著提升了0.33个标准差,可见,干预有效缩小了城乡学生的信息技术鸿沟。更为重要的是,干预组学生使用电脑中内置的计算机辅助学习软件后,数学学业表现也大幅改善,数学标准化测试成绩提升了0.17个标准差(Mo et al., 2013b)。实验2的分析结果与实验1类似,CAL显著提高了农村学生数学成绩0.15个标准差(Lai et al., 2015a)。

### 研究局限

研究人员通过开展国内首个关注弱势学生群体的计算机辅助学习干预研究项目,证明了计算机辅助学习对于打工子弟学校学生学业表现的积极作用。但由于实验中的样本局限于北京打工子弟学校学生,计算机辅助学习对于其他农村地区的贫困学生是否有显著效果尚未可知,这仍需要实验数据来回答。

### 研究二:“计算机辅助学习”干预在农村地区贫困学生中的有效性验证

#### 研究简介

为验证计算机辅助学习干预对农村地区学生的学业表现是否有影响,研究团队将关注点转向因父母在遥远的城市工作而选择一周都住在学校宿舍里的寄宿制学生。研究团队以陕西省农村地区的学校为研究对象(实验3),总样本包括72所农村学校的2726名三年级和五年级寄宿学生。研究团队将他们随机分为两组,对干预组开展每周2节40分钟的数学单机版CAL课程,同样在校内课外时间实施(Lai et al., 2013)。

### 研究结果

从实验3的样本数据分析结果可知,计算机辅助学习干预对农村寄宿制学生学业成绩有显著正向促进作用,相对于对照组,计算机辅助学习干预提升了干预组学生数学成绩0.12个标准差,且家庭条件较差的贫困家庭学生受干预影响更显著(Lai et al., 2013)。

### 研究局限

当然,陕西农村学校不能代表中国的所有农村学校,寄宿学生也不能代表所有农村学生。虽然研究团队以课后补充的形式验证了计算机辅助学习对提高农村学生教育产出的有效性,但由于样本集中在陕西地区,其外部有效性仍有待进一步验证。

### 研究三:“计算机辅助学习”干预在民族地区农村学生中的有效性验证

#### 研究简介

少数民族学生在学习方面可能面临着一个巨大障碍:汉语水平相对较低,这将会影响他们准确获取所学科目的教材知识。此前也有研究发现中国西北地区少数民族学生学习成绩的平均水平显著低于西北地区的汉族农村学生(Lai et al., 2015b)。研究者在实验4中,试图利用计算机辅助学习来帮助民族地区学生在课余时间提高汉语水平。研究者在青海少数民族地区选取了57所学校的1800多名三年级学生作为研究对象,为干预组(26所学校764名学生)提供每周2节40分钟语文(汉语)单机版CAL课程(均在校内课后时间实施);对照组(31所学校1125名学生)不做任何干预。

### 研究结果

研究发现CAL对少数民族学生汉语标准化考试成绩有0.14—0.20个标准差的正向影响。更重要的是,虽然研究团队只进行了语文科目的计算机辅助学习干预,但由于语文科目的干预产生的溢出效

应,研究人员也观察到了接受干预的学生数学考试分数的显著提升,平均来看他们数学成绩提升了0.22个标准差(Lai et al., 2015b)。研究团队分析,其主要原因是样本学生群体中少数民族学生比例偏高(以藏族为主,汉族占少一部分),计算机辅助学习语文干预提高了少数民族学生的语文知识水平和认汉字能力,也相应提升了其对数学知识的理解和掌握。

### 研究局限

结合前期研究,研究团队证明了CAL是一种改善学生学业表现的有效手段,特别是CAL干预在提升语言技能方面的效果说明其有潜力提高其他发展中国家语言能力较弱的学生的教育产出。但是,需要进一步思考的是,若在全国大规模推广计算机辅助学习干预项目,是否可以仍然如上述研究实验那样在校内课外时间来开展?各学校是否有足够的课外时间让学生使用计算机辅助学习软件进行补习教学?这些问题仍需开展进一步研究来回答。

## 研究四:“计算机辅助学习”干预在校内课外和课中时间开展的差异性研究

### 研究简介

考虑到CAL干预对改善我国农村学生学业表现的正向影响,如果教育部门将其扩展到大批学校中,CAL干预就有可能被纳入正常上课时间。之前的研究均在校内课外时间开展,研究人员不清楚校内课中CAL干预是否会像课外CAL干预一样有助于提高学生的学业成绩。为了回答CAL干预在校内课后时间开展与校内课中时间开展其影响是否会存在差异这一问题,研究团队在陕西选取了72所农村学校5267名三年级和五年级学生,给干预组学生提供每周2节40分钟数学单机版CAL课程,并占用学校自身的信息技术课时间(Mo et al., 2014a)。

### 研究结果

结果显示,校内课中CAL干预能够显著改善学生的学业成绩:每周2次40分钟的CAL数学课程能使三年级学生的标准化数学测试成绩提高0.17个标准差,五年级学生的标准化数学测试成绩提高0.17个标准差(Mo et al., 2014a)。

### 研究局限

考虑到存在替代效应的可能,校内课中CAL干预仍然可以提高学生的学业成绩。该研究结果与Linden(2008)的发现不同,原因可能是教师缺勤率较低,CAL项目与学生课堂学习进度一致,替代效应被最小化了(Bai et al., 2016)。将干预科目的计算机辅助学习放在信息技术课上完成,面临的挑战是学生信息技术课本应有的知识获取会被计算机辅助学习挤占或替代,其中存在的可能风险或问题是计算机辅助学习的效果是以牺牲信息技术课上课时间为代价的。

## 研究五:“在线计算机辅助学习”与“单机版计算机辅助学习”的影响差异研究

### 研究简介

随着互联网的飞速发展和“校校通”“班班通”政策的落地,95%的农村学校都接入了高速的互联网宽带。在现代信息技术日益发展的背景下,在线计算机辅助学习(Online Computer Assisted Learning,简称OCAL)是否也能够对农村学生的学业表现产生影响?如果有影响,与单机版相比,在线计算机辅助学习又有何优势?根据以往研究经验,OCAL可能有以下优势:首先,不用在每台计算机硬件上手动安装和维护软件,这有利于降低开展CAL研究的成本。其次,随时随地地登录CAL会增加学生使用软件的频率和机会。再次,OCAL软件系统允许我们将社交功能整合到学生用户的交互中,例如,学生可以在测验中与同龄人竞争并获得虚拟奖品,该系统还可以提供排行榜功能,以进一步激励学生学习。

为验证“在线计算机辅助学习”干预改善农村学生学业表现的有效性,研究团队在44所农村学校(陕西省31所,其他几省13所)开展了OCAL的随机干预实验(实验7),为干预组(22所学校,714名学生)提供每周2次每次40分钟的英语OCAL课程,在校内课内时间进行;对照组(22所学校,936名学生)不做任何干预(Bai et al., 2018)。

## 研究结果

研究发现,干预组学生的整体英语成绩比对照组提高了0.56个标准差,这显然比单机版CAL的影响要大。研究人员也汇总了7项随机干预实验的结果(见图6),发现计算机辅助学习干预项目对学生的学业成绩(语文、数学和英语)都产生了正向的提升效果,显著提高了农村学生的教育产出(Bai et al., 2018)。

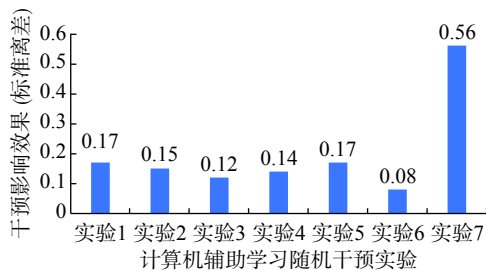


图6 计算机辅助学习对农村学生学业表现的影响

关于计算机辅助学习干预的影响机制,研究人员在排除了霍桑效应或自发变化的可能性之后,认为兴趣激发是学生学习成绩提高的主要原因。此外,由于计算机辅助学习软件设计的趣味性,受干预的农村学生有机会与同龄人进行比较和竞争,同时,研究人员还针对每个学生的个人需求为其量身定制了补习题库,这些都可能有助于样本学生学习成绩的稳步提高(Bai et al., 2018)。

## 研究局限

该研究是中国第一个关于在线计算机辅助学习的干预研究。研究人员认为,干预科目英语是农村地区学生普遍水平较低的科目,其起点低、提升空间大,因此可能存在高估在线计算机辅助学习效果的风险。因此,还需要继续开展更多在线计算机辅助学习干预研究,以进一步验证其影响效果,深入探讨其影响机制。

## 研究六:“计算机辅助学习”干预对农村学生非学业表现的影响

### 研究简介

衡量教育产出不应只关注学生的学业表现,也应关注其非学业表现。在以上实验中,研究团队也测量了一些能识别学生非学业表现的结果变量,如自我效能感(通过自我效能感量表进行测量)、喜欢上学的程度(学生对自己喜欢上学的程度进行打分)、喜欢相应干预科目的程度(学生对自己喜欢某门课程的程度进行打分)、喜欢老师的程度(学生对自己喜欢该门课程老师的程度进行打分)和对自己未来的教育期望(选择自己未来期待的最高受教育程度)等。在分析计算机辅助学习干预对农村学生学业表现影响的同时,研究人员也关注其是否对农村学生的非学业表现产生了影响。

## 研究结果

汇总以上7项实验结果可以发现,总体上计算机辅助学习对农村地区学生的非学业表现也显示出了正向影响,从表2的不同实验结果可知,农村学生的自我效能感显著提高(见表2,实验1和实验4),对所学课程的喜爱程度明显上升(见表2,实验2和实验7),同时自身的教育期望也显著提升(见表2,实验7)。

## 研究局限

国内学界对计算机辅助学习对学生非学业表现的研究偏少,因此,其影响机制尚未被清晰揭示,这也值得进一步研究探讨。

## 3. 小结

在现代信息技术高速发展的今天,信息技术与教育的关系也日益密切。从教育产出的角度出发,研究现代信息技术在农村教育中的应用现状及其影响,对促进教育均衡发展有很高的科学价值和实践意义。从前期研究来看,以计算机辅助学习为代表的教育信息技术能在农村学生成绩落后的情况下显著改善学业表现和非学业表现,且具有简单易行、行之有效、易于大规模推广的优点。这些教育信息技术手段在促进农村义务教育优质均衡发展和缩小城乡教育差距方面具有巨大潜力,但信息技术手段



影响教育产出的途径和作用机理仍有待深入研究。

### (三) 随机干预实验在“营养健康与教育”领域的应用

#### 1. 研究问题

教育人力资本和健康人力资本是人力资本的两个重要组成部分。通过上述几项随机干预实验在教育领域的应用研究,不难发现投资教育可以提高个体知识技能,提高其在劳动市场上的生产效率,因而可以提高人力资本质量。如果投资健康,则不仅可以增加人力资本质量,而且可以通过延长工作年限来增加人力资本数量。

要提升我国人力资本的整体质量,农村学生的健康问题理应引起关注。这是因为,从干预对象来看,越早开展的人力资本投资回报就越高(Heckman et al., 2003),也就是说,对学龄阶段儿童的健康卫生问题进行投资,及时解决相关问题,可能有更高的投资回报;从干预内容来看,对学生健康卫生问题的改善既是对健康的投资,也是对教育的投资。在我国农村贫困地区,缺铁性贫血和视力问题是农村学生面临的两类普遍存在而又值得关注的健康挑战。

#### (1) 农村学生的缺铁性贫血问题

缺铁性贫血是在世界范围内普遍存在的营养健康问题,尤其是在发展中国家。全世界约四分之一的人患有缺铁性贫血(De Benoist et al., 2008)。在中国农村地区,学生的营养健康问题也不容忽视。根据世界卫生组织定义,血红蛋白水平低于 115 g/L(适用于 11 岁及以下儿童)或 120 g/L(适用于 12 至 14 岁儿童)即为贫血。一项汇总了 2009 年至 2013 年的 27 项调研、涉及全国 10 个省份共 27535 名学生的研究表明,农村学生缺铁性贫血平均发生率为 27%(Zhou et al., 2015)。在四川、陕西、宁夏、青海、甘肃等西部地区针对小学生缺铁性贫血开展的研究结果也基本与此一致,西部农村小学生平均缺铁性贫血率在 25%—30%;90% 以上的学校缺铁性贫血发生率在 5% 以上,70% 以上的学校缺铁性贫血发生率在 10% 以上(罗仁福等, 2011; Luo et al., 2011; Luo et al., 2011a, Luo et al., 2011b; Zhang et al., 2013; Li et al., 2018)。随着农村地区“撤点并校”政策的实施,有较多学生开始选择住校。有研究表明,相对非住校学生,住校学生的缺铁性贫血问题更严重,学生在校期间的营养均衡问题更需要关注(Luo et al., 2011; Luo et al., 2011b)。当然,随着经济发展和社会福利条件的改善,很多营养健康问题在逐渐被解决,但农村贫困学生的缺铁性贫血情况并没有表现出显著降低的趋势(Luo et al., 2012a)。

农村学生缺铁性贫血问题的发生,主要是因为学生饮食不均衡,缺少对营养丰富食物的摄入,例如西部地区农村家庭的饮食以面条、大米等葡萄糖、蛋白质等含量丰富的食物为主,蔬菜、水果、红肉等含铁、维生素等微量元素丰富的食物相对摄入较少(Luo et al., 2011)。此外,学生家长和学校负责人对缺铁性贫血问题的认知水平也亟需提高(Luo et al., 2011)。

缺铁性贫血会给学生的成长、发展造成严重的负面影响。研究表明,因血红蛋白含量低,缺铁性贫血会影响血液向大脑及身体输送氧气,导致困乏、容易疲劳、注意力不集中等问题(Dallman, 1986; Iannotti et al., 2006)。较多研究表明,缺铁性贫血对学生的认知能力发展、学业表现、情绪管理等有不可逆转的影响(Lozoff et al., 2000; Luo et al., 2012a)。患有缺铁性贫血的孩子长大后收入可能更低(Halterman et al., 2001; Bobonis et al., 2006)。而且,婴儿时期或儿童早期阶段的缺铁性贫血产生的负面影响不可逆转,即使是后期缺铁性贫血得到了治疗,早期缺铁性贫血造成的负面影响仍然会在成年期表现出来(Lozoff et al., 2000; Lozoff & Georgieff, 2006)。

#### (2) 农村学生的视力问题

关注农村地区儿童视力健康是“教育精准扶贫”和“健康精准扶贫”的共同需求。近年来,中小学课内外负担不断加重,电子产品不断普及,加上学生用眼不健康、缺乏户外体育活动等因素,我国儿童青少年近视率居高不下,近视低龄化、重度化等问题日益严重,已成为一个关系国家和民族未来发展的不容忽视的问题(中华人民共和国国家卫生健康委员会, 2018)。研究人员在广东省开展的一项研究表明,近 22% 的 5—15 岁农村学生存在视力问题(He et al., 2004)。2012 年研究人员在甘肃和陕西两省

开展的面向 19934 名农村学生的调查研究显示, 25% 的四、五年级学生有视力问题(Ma et al., 2014)。一项在湖南、湖北、安徽等中部六省开展的针对 16187 名农村学生视力问题的调查显示, 小学生近视发生率为 42%(华文娟等, 2013)。

实际上大多数学生的视力问题都可以通过简单的视力监测(例如视力筛查)发现, 并通过及时配戴合适的眼镜加以矫正(He et al., 2007; Ma et al., 2014; WHO, 2014)。尽管近视矫正方法(配戴一副合适的眼镜)经济、简单且收益显著, 但中国农村学生的近视矫正率仍然不到 20%(He et al., 2007; Li et al., 2008; Congdon et al., 2008; Sylvia et al., 2018)。即使学生在经过高质量的视力筛查后得知自己存在视力问题, 仍有近 2/3 的学生没有配戴合适的眼镜(Li et al., 2010)。科学研究表明, 如果儿童时期的视力健康问题未得到及时矫正(如配戴合适的眼镜), 则可能发展为弱视, 甚至产生永久性视力损伤, 从而影响未来的学习效率和成绩(Resnikoff et al., 2008; Chadha et al., 2011; Luo et al., 2012b; Yi et al., 2015; Glewwe et al., 2016)。而配戴合适的眼镜不仅能够改善视力状况、改善生活质量, 还能够提高学生的阅读能力和学习成绩(Ma et al., 2014; WHO, 2014; Sylvia et al., 2018)。

农村地区学校的视力监测工作及基层医疗卫生服务供给不足, 是农村学生近视问题矫正率低的重要因素(白云丽等, 2015)。视力监测方面, 有超过 50% 的农村学校在最近两年未能按照教育部要求“每学期对学生进行两次视力筛查”(Bai et al., 2014)。即使组织了视力筛查, 参加视力筛查的学生中仍有近 30% 的近视学生未被检出, 且对近视学生的矫正率并没有显著影响(Sharma et al., 2008; Bai et al., 2014)。在基层医疗卫生机构服务供给方面, 县级以下医疗机构针对近视的治疗几乎是空白。38% 的县医院没有或仅有一个眼科医生, 超过 80% 的乡镇卫生院不具备提供矫正视力服务的能力, 而村级诊所则基本不提供视力保健相关服务。此外, 农村地区学生及家长对近视问题的重视程度不高, 对矫正视力问题普遍缺乏正确认知(Li et al., 2010; Yi et al., 2015)。甚至有超过 1/3 的校长、家长和学生都错误地认为“戴眼镜会导致视力越来越差”(Li et al., 2010; Bai et al., 2014; Yi et al., 2015)。

## 2. 随机干预实验在“营养健康与教育”领域的应用及研究结果

为探索解决农村学生缺铁性贫血问题和视力问题的有效手段, 国内研究者在该领域设计、实施了十多项随机干预实验。这些随机干预实验的目标主要集中在以下几个方面: 首先, 学生自身的营养健康问题是否得到改善, 干预实施后学生血红蛋白水平是否提升、缺铁性贫血率是否下降, 学生是否得到视力保护服务、近视学生配戴眼镜比例是否提高; 其次, 随着营养健康状况的改善, 学生的学业表现是否因此得到改善; 最后, 对不同类型的干预研究进行成本效益对比分析, 探讨成本更低的干预手段。

### 研究一: 随机干预实验在解决农村学生缺铁性贫血问题中的应用

为探索解决农村学生缺铁性贫血问题的可行方案, 自 2008 年至 2012 年, 研究团队先后开展了 6 次大规模随机干预实验研究项目, 总样本涉及 489 所农村小学、约 2 万名小学生(见表 3)。这 6 项随机干预实验研究着眼于多层面并着力探索多种干预方式。从实施层面看, 项目涉及学生本人、家长、学校负责人、地方政府部门及社会力量等。从干预方式看, 主要可以分为三大类: (1) 添加铁补充剂(例如含铁的多维元素片); (2) 以提升家长或学校负责人贫血相关认知为目标的信息干预; (3) 以改善学校管理为目标为学校补贴和激励干预(Luo et al., 2012a; Luo et al., 2012b; Miller et al., 2012; 常芳等, 2013; 史耀疆等, 2013; Kleiman-Weiner et al., 2013; Sylvia et al., 2013; Wong 2014; Zhang et al., 2013; Mo et al., 2014b; Luo et al., 2019a)。这些随机干预实验项目的抽样方式及调研形式略有不同, 但都确保了科学开展随机干预实验的要素: (1) 样本代表性, 在县级层面随机抽取农村小学的四、五年级(8 至 12 岁)学生; (2) 避免样本污染, 随机化均在学校层面实施而非学生层面, 以避免同校内不同干预组学生相互交流污染实验。

#### (1) “添加铁补充剂”干预

如前文所述, 西部农村贫困地区学生的缺铁性贫血主要是由饮食不均衡、微量元素铁等摄入不足引起的, 因此, 设计随机干预实验可以从这个角度入手。有研究发现, 当出现缺铁性贫血后, 补充铁等

微量元素(例如服用包括铁、维生素等的多维元素片)是最直接且常见的做法(Luo et al., 2012a; 史耀疆等, 2013)。那么,这一操作简单、并且易于监控的方法是否可以解决农村学生因饮食不均衡而出现的缺铁性贫血问题? 如果有效,是否可以在农村学校进行大规模推广? 这个问题通过以下随机干预实验研究来回答。

### 探索1: 每天补充一片多维元素片能改善农村学生的缺铁性贫血状况吗?

自2008年至2010年,为检验补充多维元素片能否改善农村学生的缺铁性贫血情况并提升学生的学业表现,研究团队在陕西省120多所农村学校开展了两次随机干预实验研究(见表3,实验1和实验2)。干预的具体做法是:学生在校期间,研究人员给班主任分发包含21种微量元素(其中铁元素的含量为5 mg)的多维元素片,请班主任每天在自己办公室烧好开水,然后将学生集中到班主任办公室,让学生每人用温水服用一片多维元素片。周末则请班主任给每名学生发放2片多维元素片,让学生带回家服用。研究项目执行期间,干预组的每名班主任每月可获得100元的补贴(Luo et al., 2012a; Wong et al., 2014)。

表3 随机干预实验在农村学生缺铁性贫血问题中的应用研究汇总表

序号	时间	地点	干预方式及样本量	干预效果	参考文献
1	200810-200906	陕西	样本: 共66所小学, 3661名学生 对照组: 30所学校1607名学生 干预组1: 24所学校1413名学生 干预内容: 每天补充一片含微量元素的多元维生素片(热水送服) 干预组2: 12所学校641名学生 干预内容: 向家长发放《告家长书》, 告知家长学生是否贫血、贫血的影响及如何避免贫血	相对于对照组: 干预组1(维生素组)血红蛋白水平显著提高2.14 g/L (0.2SD), 标准化数学成绩显著提高0.20SD; 干预组2(家长信息组)对血红蛋白水平及标准化数学成绩均无显著影响	Luo et al., 2012a
2	200911-201006	陕西	样本: 共60所小学, 2249名学生 对照组: 30所学校1051名学生 干预组1: 15所学校674名学生 干预内容: 每天补充一片含微量元素的多元维生素片(热水送服) 干预组2: 15所学校524名学生 干预内容: 对家长进行一次面对面的培训, 约2个小时	相对于对照组: 干预组1(维生素组)学生血红蛋白水平显著提高3.69 g/L (0.33SD), 标准化数学成绩显著提高0.1SD; 干预组2(家长培训组)学生血红蛋白水平提升2.22 g/L (0.2SD), 标准化数学成绩无显著影响	Wong et al., 2014
3	201011-201106	甘肃	样本: 共70所小学, 2686名学生 对照组: 30所学校1096名学生 干预组1: 20所学校757名学生 干预内容: 每天一片可咀嚼的含微量元素的维生素片 干预组2: 20所学校833名学生 干预内容: 每天一枚鸡蛋	相对于对照组: 干预组1(咀嚼维生素组)血红蛋白水平显著提高2.6 g/L (0.2SD), 标准化数学成绩显著提高0.12SD; 干预组2(鸡蛋组)对血红蛋白水平和标准化数学成绩均无显著影响	Kleiman-Weiner et al., 2013
4	200909-201005	青海、宁夏	样本: 共72所小学, 3553名学生 对照组: 27所学校1623名学生 干预组1: 15所学校596名学生 干预内容: 对校长进行一次贫血相关知识培训 干预组2: 15所学校667名学生 干预内容: 对校长进行一次贫血相关知识培训, 并按照1.5元/学生/天给予学校财政补贴 干预组3: 15所学校667名学生 干预内容: 对校长进行一次贫血相关知识培训, 并按照1.5元/学生/天给予学校财政补贴, 同时针对贫血学生的减少率对校长进行现金激励, 减少一个贫血学生奖励约150元	相对于对照组: 干预组1(信息组)和干预组2(补贴组)对血红蛋白水平无显著影响; 干预组3(激励组)血红蛋白水平显著提高2.4 g/L, 缺铁性贫血率降低了5个百分点	Miller et al., 2012



续表 3

序号	时间	地点	干预方式及样本量	干预效果	参考文献
5	201109-201205	青海、甘肃和陕西省	样本: 共170所小学, 干预周期为8个月 对照组: 32所学校 干预内容: 按照0.3元/学生/天进行补贴(小额补贴), 无校长激励 干预组1: 33所学校 干预内容: 按照0.7元/学生/天进行补贴(大额补贴), 无校长激励 干预组2: 20所学校 干预内容: 减少一个贫血学生激励12.5元+0.3元/学生/天补贴 干预组3: 20所学校 干预内容: 减少一个贫血学生激励12.5元+0.7元/学生/天补贴 干预组4: 33所学校 干预内容: 减少一个贫血学生激励125元+0.3元/学生/天补贴 干预组5: 32所学校 干预内容: 减少一个贫血学生激励125元+0.7元/学生/天补贴	相对于对照组(无激励+小额补贴): 干预组1(无激励+大额补贴)贫血率显著下降14.5个百分点; 干预组2(小额激励+小额补贴)对贫血率无显著影响; 干预组3(小额激励+大额补贴)对贫血率无显著影响; 干预组4(大额激励+小额补贴)贫血率显著下降13.8个百分点; 干预组5(大额激励+大额补贴)贫血率显著提高19.6个百分点	Luo et al., 2019a
6	2011-2012	宁夏	样本: 共51所小学, 900名学生, 干预周期为1年 对照组: 450名学生 干预组1: 225名学生, 学生家长每周收到贫血相关知识的短信 干预组2: 225名学生, 学生家长每周收到贫血相关知识的短信, 并且每月会对家长进行贫血相关知识的测试, 给予10元左右的现金奖励	相对于对照组: 干预组1(信息组)对家长相关知识水平、学生血红蛋白水平、学生健康水平、标准化数学成绩、上课注意力集中程度没有显著影响; 干预组2(信息+激励)家长相关知识水平显著提高了0.23SD, 学生健康水平显著提高了0.15SD, 学生数学成绩显著提高了0.14SD, 学生上课注意力集中程度显著提高了0.19分, 但学生血红蛋白水平没有显著提高	Mo et al., 2014b

研究结果表明, 每天给学生补充一片多维元素片, 补充 7 个月后可以显著提高学生的血红蛋白水平 1.7—2.3 g/L(约相当于 0.2 个标准差); 贫血状况的改善也带来了学生学业表现的提高, 学生的标准化数学测试成绩显著提高了约 0.1—0.2 个标准差(Luo et al., 2012; Wong et al., 2014; 史耀疆等, 2013)。此外, 这一干预对学生的心理健康状况也产生了积极影响, 研究团队利用心理健康诊断测验(MHT)测量学生心理焦虑水平, 结果发现测试得分显著下降了 0.3 个标准差, 这也为改善学生心理健康状况提供了除心理健康辅导外的另一种可能有效的干预手段(Zhang et al., 2013)。

研究团队做了进一步分析, 发现该干预方式虽然效果较好, 但面临一个操作上的挑战, 即每天需要班主任烧开水再组织全班学生服用多维元素片, 这在一定程度上增加了班主任的工作负担。虽然研究人员为班主任提供了每月 100 元的额外补贴(相当于约 1—2 天的工资), 但组织学生服用多维元素片作为一项额外的工作在农村学校大规模推广时仍会面临一定阻力。因此, 研究团队进一步探索了操作更方便的干预方式。

**探索 2: 服用可咀嚼的多维元素片对改善农村学生缺铁性贫血效果更好吗?**

如果上述研究在操作层面存在的挑战可能影响农村学生服用多维元素片的依从率, 那么将需要热水服用的多维元素片改为可直接咀嚼的多维元素片, 干预效果是否会更好呢? 一方面, 这样可以减少班主任工作量, 可能会提高其参与干预研究的积极性; 但另一方面, 由于学生可以直接在教室咀嚼, 减

少了班主任对组织学生服用多维元素片的参与,也可能因此减少了班主任的监督作用,从而有可能降低学生服用多维元素片的依从率。研究团队进一步开展了随机干预实验研究来评估可咀嚼多维元素片到底会对农村学生的缺铁性贫血情况产生正面影响还是负面影响。

2010年研究团队在甘肃70所学校进行的这项随机干预实验除了将用热水送服的多维元素片替换为可咀嚼的多维元素片之外,其他干预内容均与探索1一致,即还是让干预组的学生每天在校内服用一片可咀嚼的多维元素片(铁含量同样为5mg),周末将两片多维元素片带回家(见表3,实验3,干预组1)。

结果表明,干预6个月后学生的血红蛋白水平显著提高了约2.6g/L(约相当于0.2个标准差);学生的标准化数学测试成绩也显著提高了0.12个标准差;此外,从成本角度分析,可咀嚼的多维元素片每片成本约0.4元,这表明只要多增加学生当时补贴的10%即可以解决缺铁性贫血这一问题(Kleiman-Weiner et al., 2013)。

### 探索3:鸡蛋可以替代多维元素片吗?

虽然上述随机干预实验研究已经证明了每天补充一片多维元素片对改善农村学生的缺铁性贫血状况、提高学生学业表现均有显著正面影响,但研究团队在与地方政府沟通中发现,很多地方的政府部门更热衷于推进另一种干预,即每天给学生补充一个鸡蛋(Kleiman-Weiner et al., 2013)。这可能是因为:(1)购买鸡蛋比较方便,而多维元素片在农村地区的普通市场买不到;(2)对农村家庭来说,通常认为鸡蛋比较有营养(Kleiman-Weiner et al., 2013)。事实上,鸡蛋富含蛋白质,确实比较有营养,但农村学生所缺少的微量元素(尤其是铁)含量不足,一个鸡蛋平均只含有0.5mg的铁(US Department of Agriculture, 1999)。研究团队发现地方政府在给农村学生发放鸡蛋方面的投入很大,但这能否改善农村地区学生的营养状况进而提升学生的学业表现呢?

为评估地方政府部门给农村学生发放鸡蛋的政策影响,2010年研究团队在甘肃70所学校进行了一项随机干预实验(见表3,实验3),干预内容为班主任每天给学生提供一个煮好的鸡蛋(班主任因该工作每周可以得到50元的补贴)。

研究结果表明,每天补充一个鸡蛋连续干预6个月对学生的贫血状况没有显著影响,对学生的成绩也没有显著影响。这是否因为家长知道学生在学校多吃了一个鸡蛋而减少了在家吃饭时的营养供给,即产生了“挤出效应”?通过对学生是否住校的异质性分析显示,没有证据表明存在这种挤出效应(Kleiman-Weiner et al., 2013)。不难看出,发放鸡蛋这一干预不仅没有取得预期效果,而且相对于补充可咀嚼的多维元素片成本更高(每个鸡蛋约0.7元,而每片多维元素片约0.4元)。当时这一干预项目投入巨大,如果要在农村学校继续实施并推广,需要研究人员和政府部门认真考虑提高其成本产出比(Kleiman-Weiner et al., 2013)。

### (2)信息干预

缺少信息来源和相关知识储备不足一直被认为是不及时应对健康问题的重要原因(Cochrane et al., 1982; 罗仁福等, 2013; Black et al., 2013)。如果能够改善家长或学校负责人关于缺铁性贫血的相关知识储备和认知,他们就有可能主动去改善学生饮食结构以保证其摄入均衡营养,从而在主观上避免缺铁性贫血发生的风险,这可能是更长远地解决学生营养问题的办法(Wong et al., 2014)。但如何才能改善家长或学校负责人的相关知识和认知呢?这类健康教育应该瞄准家长还是学校负责人呢?研究团队做了一系列探索来回答这些问题。

### 探索1:发放一封《告家长书》对改善农村学生缺铁性贫血是否有效果?

一般认为,相对于学校,家长作为学生最主要的监护人更有动力去改善学生的营养健康状况(Mo et al., 2014b)。所以研究团队最先尝试了对家长进行营养健康知识教育的信息干预,主要包括以下信息:缺铁性贫血的基本介绍,包括特点、原因等;孩子的缺铁性贫血情况;缺铁性贫血问题的应对办法。

为评估提供这些信息是否有效果,研究团队2008年在陕西省66所学校开展了一项随机干预实验研究(见表3,实验1),干预内容是:检查完农村学生的贫血状况后,由学生带给家长一封《告家长书》,班主任会帮助检查学生是否将《告家长书》带给了家长。这封《告家长书》的内容主要包括如下几方面:(1)什么是缺铁性贫血及其影响;(2)避免缺铁性贫血有哪些可行办法,比如平衡饮食、多吃富含维生素C的水果和蔬菜、服用铁补充剂等;(3)告知家长孩子现在是否患有缺铁性贫血(程度包括:严重贫血、中度贫血、轻度贫血及不贫血);(4)对于患有缺铁性贫血的学生,会告知家长需要带孩子到医院做进一步检查,平衡饮食对改善缺铁性贫血很重要。

结果表明,平均来看,发放《告家长书》这种干预对农村学生的血红蛋白水平及学业表现均没有显著影响;异质性分析表明,该干预对非住校生的血红蛋白水平有较小但显著的提升(提升了1.30 g/L,约相当于0.1个标准差),但对其标准化数学测试成绩仍没有显著影响。研究团队分析认为只通过发放《告家长书》给家长提供营养健康知识的信息干预并不足以改变家长的行为,还需开展下一步实验研究来探索提高农村学生营养健康水平的有效方式(Luo et al., 2012b)。

### 探索2:对家长开展“面对面培训”是否能改善农村学生缺铁性贫血?

只向家长发放《告家长书》未发挥作用,是否因为信息干预的强度不够呢?开展“面对面”的家长培训和交流会会不会有效果呢?研究团队2009年在陕西省60所学校开展了一项随机干预实验研究(见表3,实验2),干预内容如下:由西安交通大学医学院的营养师在干预组学校开展家长营养健康知识培训,要求每名家长至少有一名家长参加。家长培训时长约1小时,通过幻灯片演讲、视频播放、营养知识手册讲解(可以带回家)、自由提问等形式向家长讲述以下信息:(1)缺铁性贫血的基本介绍及主要症状;(2)农村地区缺铁性贫血问题的高发现状(研究人员并未告知家长自己孩子的贫血情况);(3)缺铁性贫血对学生健康和学习的负面影响;(4)应对缺铁性贫血问题的办法(平衡饮食、多吃铁含量丰富的食物等)。

研究表明,家长培训干预实施6个月后学生的血红蛋白水平显著提升2.22 g/L(约相当于0.2个标准差),但学生标准化数学测试成绩无显著变化;异质性分析也表明该干预将非住校生的血红蛋白水平显著提高了3.27 g/L,但对其标准化数学测试成绩仍然没有显著影响(Wong et al., 2014)。这项随机干预实验研究结果表明,相对于仅发放《告家长书》,有交流和互动的面对面家长培训在提升农村学生营养健康状况方面更有效,但对提升学生学业表现仍无显著影响。

### 探索3:更准确地瞄准干预对象是否能改善农村学生缺铁性贫血?

结合上述家长信息干预研究的效果来看,改善家长的营养健康知识和认知对学生血红蛋白水平的改善效果有限,而对学生标准化数学测试成绩没有显著影响。是不是还需再瞄准干预对象才能准确验证效果呢?数据显示,随着农村小学撤点并校工作的推进,至2010年农村小学生约三分之一选择住校(Miller et al., 2012)。鉴于学校在学生饮食中发挥着越来越重要的作用,是否可以通过在学校层面实施干预(比如对学校负责人进行信息干预)来减少学生缺铁性贫血的发生?以往研究表明,在学校层面开展干预成本相对更低,但由于学校负责人较忙,其改善学生营养健康状况的动力相对更弱,这就需要他们在教学管理事务和学生健康之间做出平衡(Sylvia et al., 2013)。

为评估对学校校长开展信息干预的影响效果,研究团队2009年在青海和宁夏的72所学校开展了一项随机干预实验(见表3,实验4),对干预组学校校长进行一次约2小时的营养健康知识培训。培训内容包括:(1)告知校长该校患有缺铁性贫血学生的比例;(2)解决贫血问题的重要性,贫血对学习成绩、认知能力发展等有负面影响;(3)降低缺铁性贫血发生率的方法(补充多维元素片、在食物中增加红肉供给、对家长进行贫血相关知识教育等)。

研究结果表明,对校长进行贫血知识相关的信息干预7个月后并未显著改善学生的血红蛋白水平。研究团队后续也验证了校长贫血知识信息干预效果的异质性影响,发现当学校本身存在以学生成



绩为导向的激励考核时,该干预有显著的影响效果。研究人员分析,可能的原因是:校长可支配的资源相对有限(例如教育部门拨付给学校的生均经费是相对固定的),相对于学生的营养健康,校长可能更优先关注学生的学习成绩,将有限的资源用于提高学生的成绩(Miller et al., 2012; Sylvia et al., 2013)。

#### **探索 4: 发送手机短信干预是否能更便宜、更有效地改善农村学生缺铁性贫血状况?**

研究团队在 2011 年开展前期研究时发现,90% 的农村家庭都会使用手机。由于上述信息干预研究均未在改善学生缺铁性贫血方面取得较好效果,研究团队在考虑手机短信干预成为一种有效干预方式的可能性。与发放《告家长书》和开展家长培训相比,手机短信信息干预可能存在一些优势:(1)能够接触到更多学生家长(可能有学生家长不参加面对面培训);(2)面对面家长培训难以多次组织,而发短信则可以不限次数;(3)相对于组织培训会,发短信也更便捷、更便宜(Mo et al., 2014b)。

2011 年研究团队在宁夏 51 所小学开展了手机短信干预的随机实验研究项目以评估其影响(见表 3, 实验 6)。这项实验开展了两类干预。第一种干预方式为仅发送短信(称为“短信组”)。基线调研时研究团队收集了每名学生家长的联系方式,每周给家长发送一条关于贫血及健康方面的短信,干预一年间共发送 52 次短信,其中第一次短信介绍了项目背景,其余 51 次短信内容不尽相同,内容简短易懂,涵盖以下主题:(1)什么是缺铁性贫血和样本县的缺铁性贫血发生率;(2)缺铁性贫血问题不及时解决对成绩的负面影响;(3)如何通过均衡饮食避免缺铁性贫血的发生,比如在家庭日常饮食中增加孩子的营养摄入(建议家长多吃一些富含铁的红肉和富含维生素 C 的水果蔬菜等)和补充多维元素片等。

第二种干预方式为发送短信加知识测试(称为“短信测试组”)。除每周给家长发送与“短信组”内容一样的短信外,每月额外增加一次知识测试。知识测试方式为多选题,测试内容为本月家长收到的短信信息,研究人员会将家长未答对题目的正确答案发送给家长;若家长知识测试回答正确可获得 10 元的电话费充值奖励(约相当于当时农村家庭平均收入的 2.5%),这个奖励金额足够吸引家长注意,同时又不致于因金额过高而产生收入效应(Mo et al., 2014b)。

研究结果表明,“短信组”对家长的营养健康相关知识水平、学生的血红蛋白水平、身体健康水平、上课注意力集中程度和标准化数学测试成绩均没有显著影响。“短信测试组”家长的营养健康相关知识水平显著提高了 0.23 个标准差,学生身体健康水平显著提高 0.15 个标准差,上课注意力集中程度显著提高了 0.19 个标准差,标准化数学测试成绩显著提高了 0.14 个标准差,但学生血红蛋白水平没有显著改善。异质性分析结果表明,由于 80% 的学生家庭是由妈妈负责做饭和购买食材,因此妈妈接收短信的干预效果更好(Mo et al., 2014b)。

综合分析上述几次信息干预探索研究,不难看出,试图通过信息干预改善农村学生缺铁性贫血状况,其影响效果并不乐观,仅向家长或校长提供营养健康知识不足以改变其行为,而增加相应资源或激励后效果更佳。

#### **(3) “补贴和激励”干预**

以往研究表明,要从学校层面改善农村地区学生的缺铁性贫血情况,只提高校长关于贫血的相关知识是不够的,可能还需要其他激励。(1)相关资源(例如补贴)。由于既有资源相对较固定,想要校长额外关注学生营养健康状况就需要有相应额外资源投入。(2)激励机制。资源有限的前提下,校长需要平衡学生成绩与营养健康状况,增加一定的激励机制可能会引导他们更多关注学生的营养健康。补贴或激励更能发挥校长的积极性,使其能主动寻找方法来改善学生缺铁性贫血状况、平衡各类教育资源(Miller et al., 2012; Sylvia et al., 2013; Luo et al., 2019a)。

#### **探索 1: 提供学校补贴和校长激励能否改善学生缺铁性贫血状况?**

为验证提供学校补贴和校长激励对改善学生缺铁性贫血的影响,研究团队 2009 年在青海和宁夏 72 所农村小学开展了一项随机干预实验项目(见表 3, 实验 4)。“学校补贴组”干预内容为:先向所有校长提供与上述信息干预实验研究相同的缺铁性贫血相关知识培训,再向学校额外提供一笔每生每天 1.5 元的补贴(以当时物价计算,该笔补贴可以买 75 克红肉),让校长将补贴用于改善学生的缺铁性

贫血状况。实际操作中研究人员不监督校长如何使用这笔补贴金,换言之,校长对使用该笔补贴费用有较大自主权(Miller et al., 2012; Sylvia et al., 2013)。

“学校补贴+校长激励组”的干预内容为:除与“学校补贴组”一致的信息干预和学校补贴外,研究人员会对比基线调研时该校学生患缺铁性贫血的比例,评估调研时该校每减少一名患缺铁性贫血的学生,校长将获得150元的奖励(根据当地教师月平均工资2500元来看,如果校长能将全校学生贫血率降低50%,则可以获得相当于约两个月工资的奖励);若患有缺铁性贫血的学生比例增加,校长也不会受到任何惩罚。

结果表明,干预实施7个月后,“学校补贴+校长激励组”的学生血红蛋白水平显著提升了2.4 g/L,学生缺铁性贫血率降低了5个百分点,而“学校补贴组”对学生血红蛋白水平没有显著影响。研究还发现,该干预效果具有异质性,当学校本身存在以学生成绩为导向的激励考核时,会显著增强健康激励的效果,校长激励组学生的血红蛋白水平会多提高8.6 g/L(Miller et al., 2012)。

虽然该研究在改善学生缺铁性贫血方面取得了正向影响,但仍面临一些挑战:给校长激励金额多少是最合适的?“学校补贴组”干预没有效果是因为补贴金额太少吗?增加补贴金额是否有助于改善学生的贫血状况?这都需要研究团队进行下一步探索。

### 探索2:“学校补贴多少、校长激励多少”更有助于改善学生缺铁性贫血状况?

为验证不同的学校补贴额度、校长激励额度的影响效果,2011年研究团队在青海、甘肃和陕西3省的130所学校开展了一项随机干预实验研究项目(表3,实验5)。实验执行中,除了对所有学校校长进行与前期研究相同的营养知识信息干预外,额外增加了“校长激励”和“学校补贴”两组干预,具体干预内容见下:

对校长激励组。研究团队在未告知校长该校哪些学生患有缺铁性贫血的前提下,与校长签订了激励合同,明确告诉校长研究人员会依据一学年内该校贫血学生减少的数量对校长个人进行奖励。为评估不同激励额度的影响效果,研究人员将激励额度分为“大激励”和“小激励”两种:“大激励”指每减少一名贫血学生校长可以得到125元,平均来看该组校长得到的奖励金额约相当于两个月的基本工资;“小激励”指每减少一名贫血学生校长可以得到12.5元的奖励,平均来看该组校长得到的奖励金额约相当于0.2个月的基本工资。

对学校补贴组。研究团队为学校提供一笔额外的资金补贴,分两次发放(一次是在研究开始时,另一次是在研究进行中期),且不指定用途,校长可以用这些补贴改善学生饮食,也可以用于改进校长认为重要的其他事项(例如教学)。为评估不同补贴额度的影响效果,研究人员将补贴分为“大补贴”和“小补贴”两组:“小补贴”指对学校每生每天补贴0.3元(根据前期研究测算,每天补充一片多维元素片需要约0.3元),该组学校平均获得了7452元;“大补贴”指每生每天补贴0.7元,该组学校平均获得了17388元。

研究团队干预分配时,使用了3×2交叉设计的方式,将170所学校随机分为6组:(1)“无激励+小补贴”组(即对照组),32所学校;(2)“无激励+大补贴”组,33所学校;(3)“小激励+小补贴”组,20所学校;(4)“小激励+大补贴”组,20所学校;(5)“大激励+小补贴”组,33所学校;(6)“大激励+大补贴”组,32所学校。

结果表明,干预实施8个月后,对校长实施大额激励干预将学生的缺铁性贫血发生率显著降低了13.8个百分点,约相当于基线测试时缺铁性贫血发生率的37.9%,给予校长小额激励对学生缺铁性贫血状况没有显著影响。对学校提供大额补贴同样将学生缺铁性贫血发生率显著降低了14.5个百分点,约相当于基线缺铁性贫血率的39.8%。但在“大激励+大补贴”组,与对照组相比,学生缺铁性贫血情况不但没有改善,贫血发生率反而显著提高了19.6个百分点。研究团队认为,校长激励和对学校提供补贴对校长产生了两种不同的影响,两者相互替代而非相互补充,给校长提供大额激励会挤出提供大额补

贴的影响,因此出现了“大激励+大补贴”反而对改善学生缺铁性贫血产生了负面影响。

综合提供校长激励和学校补贴相关随机干预实验研究可知,这两种干预均可显著改善学生缺铁性贫血情况,但从成本效益分析来看,“校长激励”方式每减少一名缺铁性贫血学生花费的成本更低(平均 723 元),约为“学校补贴”方式每减少一名缺铁性贫血学生花费的成本的一半(平均 1447 元)(Luo et al., 2019a)。

## 研究二:随机干预实验改善农村学生视力健康状况的应用研究

针对上文提到的中国农村地区学生近视发生率高、近视矫正率低的问题,为积极响应“教育精准扶贫”战略,探索改善农村学生视力问题、完善农村地区青少年视力健康保障体系的有效模式,研究者通过大规模实证研究了解了农村地区青少年视力健康发展面临的一系列挑战,探索了一系列视力健康方面的可行干预措施。

首先,研究者从制约农村学生视力健康寻求行为的需求和供给两方面开始探索。研究者对近视学生及家庭提供了视力健康知识信息干预,希望通过提供正确的视力保护和戴镜知识来改善学生的视力健康寻求行为,提高近视学生的眼镜配戴率,结果发现单纯提供信息干预对改善学生视力健康寻求行为并无显著影响。同时,研究者也设计了在学校为近视学生发放配镜补贴(含免费眼镜和眼镜兑换券)的干预方式,结果显示提供配镜补贴干预可以显著改善学生的视力健康、学业表现和心理健康状况。

接下来,为进一步提高农村地区近视学生的眼镜使用率,研究者设计了激励农村学校教师的干预方案,研究结果显示教师激励干预能够显著改善近视学生的眼镜配戴率和使用率。

为探索高效、可持续的解决农村地区学生视力健康问题的有效模式,研究者进一步探索了与县级政府等多方合作建立县级农村学生视力保护服务中心(简称“视光中心”)的随机干预实验研究项目,评估结果显示视光中心干预能够显著改善近视学生的视力矫正率和学业表现,且能为当地居民和农村学生提供可持续的视力健康服务。

最后,直接为近视学生提供免费眼镜所面临的挑战是:学生因为免费得到眼镜而不坚持戴眼镜,或者不戴眼镜,从而出现免费眼镜使用率低、造成资源浪费的问题。为在提高近视学生眼镜配戴率的同时能够降低干预项目的推广成本,以让更多其他地区的农村学生能享受到高质量的视力健康服务,研究者采用了“费用均摊”(发放眼镜兑换券和部分配镜补贴)的方式代替直接发放免费眼镜的方式,实验结果显示“费用均摊”的干预方式提高了农村学生长期配戴眼镜的可能性。以下为四个研究项目的具体介绍(见表 4)。

表 4 随机干预实验在改善农村地区学生视力健康和学业表现中的应用研究汇总表

序号	时间	地点	干预方式及样本量	干预效果	文章来源
1	201209-201306	陕西、甘肃	<p><b>样本:</b> 共252所小学, 3177名学生, 干预持续时间9个月</p> <p><b>对照组:</b> 84所学校, 1036名学生</p> <p><b>干预组1:</b> 84所学校, 988名学生</p> <p><b>干预内容:</b> 为视力不达标学生发放配镜补贴(免费眼镜)</p> <p><b>干预组2:</b> 84所学校, 1153名学生</p> <p><b>干预内容:</b> 为视力不达标学生发放免费眼镜券, 学生家长可带学生到当地县医院兑换一副眼镜(眼镜兑换券)</p>	<p>相对于对照组:</p> <p>干预组1(免费眼镜): 近视学生配戴眼镜的可能性显著提高1.54倍, 标准化数学成绩显著提高0.11个标准差;</p> <p>干预组2(眼镜兑换券): 近视学生的配戴眼镜的可能性显著提高1.44倍, 标准化数学成绩无显著影响;</p> <p>干预组1(免费眼镜)和干预组2(眼镜兑换券): 显著降低近视学生的视力下降程度, 比对照组学生视力少下降0.23行;</p> <p>干预组1(免费眼镜)和干预组2(眼镜兑换券): 显著降低了近视学生的身体焦虑; 显著降低了学习成绩较好的近视学生的学习焦虑</p>	<p>Ma et al., 2014; Ma et al., 2015; Guan et al., 2018a</p>



续表 4

序号	时间	地点	干预方式及样本量	干预效果	文章来源
2	201309-201403	上海、苏州、无锡	样本: 共94所小学, 728名学生, 干预持续时间6个月 对照组: 47所学校370名学生 干预组: 47所学校358名学生 干预内容: 提供免费眼镜和教师激励	相对于对照组: 干预组(免费眼镜+教师激励): 近视学生配戴眼镜的可能性显著提高11.5倍; 且从之前仅提供配镜补贴41%的配镜率显著提高到68%	Yi et al., 2015
3	201409-201506	陕西	样本: 共31所小学, 949名学生, 干预持续时间9个月 对照组: 11所学校516名学生 干预组: 20所学校433名学生 干预内容: 早筛查转诊到视光中心, 提供眼镜兑换券	相对于对照组: 干预组(早筛查组): 标准化数学成绩显著提高0.25个标准差, 拥有眼镜和配戴眼镜的可能性均显著提高了1.4倍	Ma et al., 2018a
4	201409-201506	甘肃	样本: 共164所小学, 2713名学生, 干预持续时间9个月 对照组: 82所学校1461名学生 干预组: 82所学校1252名学生 干预内容: 项目开展期间将学生转诊到视光中心, 提供眼镜兑换券	相对于对照组: 干预组(眼镜兑换券): 拥有眼镜和配戴眼镜的可能性显著提高了11倍和6倍	Ma et al., 2018b
5	201409-201506	广东、云南	样本: 共127所小学, 867名学生, 干预持续时间9个月 对照组: 37所学校250名学生 干预组1: 32所学校252名学生 干预内容: 为视力不达标学生发放眼镜兑换券, 学生家长可带学生到当地县医院兑换一副眼镜(补贴100%) 干预组2: 31所学校183名学生 干预内容: 支付90元领取眼镜组(约补贴85%) 干预组3: 27所学校182名学生 干预内容: 支付180元领取眼镜组(约补贴50%)	相对于对照组: 干预组1(眼镜兑换券): 显著降低了学生长期健康寻求行为的可能性。 干预组2(支付90元领取眼镜)和干预组3(支付180元领取眼镜): 一定程度改善了学生持续寻求视力健康的行为	Wang et al., 2017

(1)信息干预和配镜补贴对学生学业表现和心理健康的影响研究

研究简介

研究团队在农村地区进行了大量的访谈,发现农村地区近视矫正率低的主要原因是信息不对称和错误观念的影响;家长、老师甚至学生自己都不知道近视,甚至还存在“近视配戴眼镜会越来越严重”的错误认知。针对这一问题,研究人员发现最常用的干预方案是提供健康信息知识或者提供免费试用品,来帮助他们了解近视及配镜相关知识,并通过试用体验到戴眼镜的好处。这样不仅能改善学生视力,还会增加他们对视力健康服务的未来需求,因此,这或许是可以长期持续改变农村学生视力健康寻求行为的有效方法(WHO, 2006; Girard et al., 2010; 李玲课题组, 2015; Dupas et al., 2014; 聂景春等, 2016; Nie et al., 2018)。然而,有研究显示,免费提供眼镜存在较高程度的浪费(Sylvia et al., 2018)。

为了提高眼镜的使用率,研究者同时设置了眼镜兑换券干预——近视学生家庭需要到县城指定地点领取免费眼镜。这种干预方式希望那些迫切需要和使用可能性大的群体,会花费时间、路费及其他经济成本去兑换眼镜,从而减少免费提供健康卫生服务和产品所带来的浪费,提高干预的效果。

因此,研究团队在陕西和甘肃两省开展了针对农村四、五年级 3177 名贫困学生的随机干预实验研究,干预设计为“提供视力健康培训”和“配镜补贴”两种,以验证提供信息干预对提高近视学生视力矫正率的影响,以及提供配镜补贴与学生视力健康、学业表现和心理健康之间的因果关系。该研究于 2012 年 9 月起在陕西和甘肃 252 所学校 3177 名四、五年级学生中开展了为期 9 个月的随机干预实验。

研究干预执行方案为:2012年9月研究团队对所有样本学校学生进行了基线调研(视力测试和标准化数学测试),随后采用3×2因子设计,将所有学校随机分成6组(见表5)。

表5 信息干预和配镜补贴随机干预实验分组情况

干预类别	免费眼镜组(84所学校)	眼镜兑换券组(84所学校)	对照组(84所学校)
信息干预(126所学校)	42所学校	42所学校	42所学校
无信息干预(126所学校)	42所学校	42所学校	42所学校

如表5所示,研究团队随机将84所学校分配到对照组(仅发放学生验光结果通知单告知家长该生的视力状况),84所学校分配到“免费眼镜干预组”(在学校给近视学生分发免费眼镜),84所学校分配到“眼镜兑换券干预组”(给近视学生发放眼镜兑换券,学生家长带学生到当地县医院可凭券兑换一副免费眼镜)。同时随机进行健康信息干预分组:一组接受关于视力知识和配戴眼镜知识的培训(即“有信息干预”),研究人员在课堂上为该组所有学生、家长和教师播放视力保护相关视频并发放视力保护健康知识培训手册,另外一组不接受培训(即“无信息干预”)。2013年6月,研究团队再次对所有学校进行了评估调研,收集与基线调研相同的信息(包括对所有学生再次进行视力测试和标准化数学测试)。

### 研究结果

视力健康信息干预评估结果显示:相比于对照组,单纯提供信息干预对提高近视学生视力矫正率没有显著的影响,而“免费眼镜+信息干预组”的视力矫正率在三周内达到了95%,这与基线调研时平均15%的近视学生视力矫正率相比,无疑是巨大的提升。研究人员分析这可能是因为仅开展信息干预并不足以促使学生快速地将知识转化为行为(Ma et al., 2014)。异质性分析表明,“免费眼镜组”的留守儿童近视矫正的可能性和配戴眼镜的可能性分别显著提高了45%和26%(Guan et al., 2018b)。

近视一旦发生就不可逆,并且随年龄增长近视学生的视力问题会越来越严重。研究团队干预9个月后分析结果发现,获得免费眼镜和眼镜兑换券的学生比对照组学生的视力少下降了视力表的0.23行(Ma et al., 2015)。不难看出,这个结果与农村学生视力保护服务需求方(学生和家长)普遍存在的“配戴眼镜会导致视力越来越差”的认识误区完全相反,配戴眼镜非但没有加速视力恶化,反而有效减缓了学生视力的下降程度。换言之,这一结果表明,对农村小学生来说,配戴合适的眼镜是解决近视矫正问题最方便有效的方法。

研究团队还发现,相对于对照组,干预组近视学生配戴眼镜9个月后标准化数学测试成绩显著提高了0.11个标准差,这表明提供免费眼镜这种操作简单的干预是有显著效果的(Ma et al., 2014)。

在农村地区,使用黑板教学是最常见的教学方式,而对于近视学生来说,由于看不清楚黑板上的板书则更可能让他们产生严重的学习焦虑(Ma et al., 2014; Guan et al., 2018a)。这项随机干预实验研究的结果也表明:总体来看,提供免费眼镜和眼镜兑换券的干预方式对改善近视学生的心理健康水平和改善学习焦虑状况影响较小且不显著,但能够显著降低近视学生的身体焦虑状况0.07个标准差。异质性分析结果显示,提供配镜补贴干预能够显著降低学习成绩较好学生的心理健康水平0.26个标准差、学习焦虑水平0.25个标准差和身体焦虑水平0.22个标准差(Guan et al., 2018a)。

### 研究局限

虽然研究团队验证了信息干预和配镜补贴对农村学生视力健康、学业表现和心理健康的正向影响效果,但实验执行过程中也存在一个挑战:农村学生配戴免费眼镜的依从率低(干预实施7个月后的追踪调研结果表明,提供免费眼镜、眼镜兑换券和发放验光结果通知单的三类学校中近视学生戴镜率分别为41%、37%和26%),而只有长期戴眼镜才会有效防治近视和改善学业表现(Ma et al., 2014)。因此提高近视学生配戴眼镜的依从率是研究人员需要进一步解决的问题。

## (2)提供配镜补贴和教师激励干预对提高近视学生眼镜配戴率的影响研究

### 研究简介

单纯提高健康卫生服务或产品的拥有率并不能带来健康收益,还需要接受者主动使用才能够实现

这一目标(Dupas et al., 2014; Sylvia et al., 2018)。提供配镜补贴干预眼镜实际使用率较低,说明只给学生提供配镜补贴是不够的(Priya et al., 2019)。那么,研究人员还可以做什么呢?结合以往研究可以发现,或许正确认知和激励机制是两个可行的路径。已有研究显示,只提供健康信息干预对改善健康产品的使用率并没有显著影响(Congdon et al., 2011; Ma et al., 2014)。有研究显示,教师与学生沟通更有效、教师容易发现学生的视力问题,也更容易组织和管理学生(张晔等, 2018)。研究团队便开始尝试通过设计激励机制干预来提升近视学生的眼镜佩戴率。不同于仅提供配镜补贴,这种激励或许更能发挥教师的监督作用,从而解决近视学生配戴眼镜依从率低的问题。结合前期研究,研究团队2013年在上海、苏州和无锡3市开展了面向94所城市打工子弟学校728名学生的“教师激励干预实验研究”。

2013年9月研究团队完成所有样本学校学生的基线调研后,随机将47所学校分配为“教师激励干预组”,其余47所为对照组。除了为干预组有视力问题的学生提供免费眼镜外,还对该校的教师进行激励(研究人员随机到学校观察自然状态下近视学生的眼镜配戴率,如果班级内近视学生配戴眼镜的人数大于等于80%,教师将得到奖励)。对对照组则不进行任何干预。2014年3月研究团队对所有样本学校进行了与基线调研内容一致的评估调研,并评估了近视学生配戴眼镜的情况。

### 研究结果

结果发现,提供教师激励干预的方式显著提高了近视学生配戴眼镜的依从率(Yi et al., 2015)。干预组近视学生配戴眼镜的可能性是对照组的11.5倍。与前期仅提供免费眼镜干预的研究中41%的眼镜配戴率相比,提供教师激励干预后近视学生配镜率显著提高到了68%。进一步的追踪结果显示,有90%的近视学生持续使用眼镜超过一个学年(Yi et al., 2015)。

### 研究局限

虽然研究团队验证了提供免费眼镜和教师激励干预对提高近视学生戴镜率的正向影响效果,但实验研究执行过程仍然面临一个挑战,即无法长期、可持续地为有视力问题的学生提供视力保护服务。由于青少年时期孩子的眼睛仍在不断发育,需要至少每6个月进行一次视力检查并及时更换合适的眼镜,一次性的配镜补贴难以让大量受视力问题困扰的农村孩子长期受益,因此探索易操作、宜推广的可持续解决农村学生视力问题的有效方式是研究者需要进一步通过实验研究来解决的问题。

### (3)“视光中心”在改善农村学生视力健康和学业表现中的影响

#### 研究简介

为探索从根源上解决农村学生视力健康问题,为他们提供长期、可持续的视力保护服务,2014年研究团队在陕西省开展了面向31所农村学校949名学生的视光中心干预研究。研究团队与地方政府、县级医院等多方合作成立了视光中心,以验证视光中心能否为当地居民和农村学生提供长期的、负担得起的视力保护服务,并评估视光中心对学生视力健康和学业表现的影响。每一个县级视光中心的工作人员都是由经过专业培训的县医院眼科医生担任,他们会培训农村教师定期对学生进行视力筛查,并由教师将筛查后视力不良的学生转诊到视光中心进行进一步验光检查并接收免费矫正(Ma et al., 2018a)。

2014年9月研究团队对所有样本学校学生进行基线调研(视力测试和标准化数学测试)后,随机将20所学校分配到“早筛查干预组”:先为该组所有学生提供视力筛查,然后为在视光中心进行了视力检查后需要视力矫正的学生提供配镜补贴。剩余11所学校分配为对照组,与干预组相比,他们被称作“晚筛查组”,即在完成“早筛查干预组”干预活动后,再为分配到“晚筛查组”的学生提供视力筛查服务及配镜补贴。2014年6月研究团队对所有样本学校学生进行了与基线调研一样的评估调研。

#### 研究结果

实证分析结果发现,“早筛查干预组”的学生近视矫正率和戴镜率均显著提高了1.4倍,学生标准化数学测试成绩显著提高0.25个标准差,这一结果表明学生近视后越早发现并及时配戴眼镜对学业成



绩的正向影响越显著(Ma et al., 2018a)。研究团队后续在甘肃也开展了类似的随机干预实验,通过县级视光中心为干预组学生提供视力筛查和配镜补贴,结果发现与对照组相比,干预组近视学生拥有眼镜和配戴眼镜的可能性均显著提高了 11 倍和 6 倍(Ma et al., 2018a)。以上结果表明,在县级地区建立以县医院为基础的视光中心,为近视学生提供配镜补贴,可能是长期、可持续地改善中国农村儿童视力健康和学业表现的有效途径。

### 研究局限

与健康卫生领域普遍面临的问题一样,农村学生的视力问题并不是短期免费试用或一次性补贴所能解决的,短期免费试用或一次性补贴干预是否具有长期效果也是该领域研究的一个重点问题(Dupas et al., 2014)。接下来,研究团队也试图验证发放配镜补贴的干预方式能否显著增加农村学生对视力健康服务的未来需求。

### (4)“费用分摊”方式补贴干预对学生视力健康寻求行为的影响研究

#### 研究简介

已有研究表明,学生因为免费得到眼镜而不坚持戴眼镜,或者不戴眼镜,而出现了免费眼镜使用率低的问题(Ma et al., 2014)。为进一步探究低成本、高效率的干预方式,研究团队分析了补贴干预与学生视力健康寻求行为变化之间的因果关系,并尝试识别“费用分摊”方式对农村学生的视力健康寻求行为的影响。结合已有研究项目,研究团队于 2014 年对云南和广东 2 省 127 所农村小学的 867 名学生进行了“补贴干预”随机干预实验研究。

2014 年 9 月,研究团队对所有样本学校学生进行了基线调研(包括基本情况调查和视力测试),之后,研究人员给视力未达标的学生发放了视力筛查结果告知单,告知家长该生可能存在视力问题并建议其带学生到视光中心进行全面视力检查和矫正。后续进行了如下随机分组:32 所学校分配到“眼镜兑换券组”(近视学生得到 100% 配镜补贴);31 所学校分配到“支付 90 元领取眼镜组”(约补贴 85%);27 所学校分配到“支付 180 元领取眼镜组”(约补贴 50%);37 所学校分配到对照组(仅发放结果告知单,无任何其他干预)。2015 年 6 月研究团队对所有样本学校学生进行了评估调研,评估了近视学生视力健康寻求行为的变化情况。

### 研究结果

实证分析结果发现,相对于对照组 23.6% 的配镜率,提供 100% 配镜补贴显著降低了学生进一步寻求视力健康行为的可能性(100% 补贴组配镜率为 12.7%),提供 85% 配镜补贴和 50% 配镜补贴干预则在一定程度改善了学生长期寻求视力健康的行为(配镜率分别为 21.3% 和 20.9%)(Wang et al., 2017)。与经过前期研究验证的提供免费眼镜改善学生眼镜拥有率相比,基于“费用分摊”的配镜补贴方式更能通过降低成本、提高眼镜配戴率的方式解决农村学生和家长对视力健康服务的低需求问题(Ma et al., 2014; Wang et al., 2017)。

### 研究局限

虽然研究团队验证了提供基于视光中心的视力筛查服务和“费用分摊”式补贴干预对提高学生配镜率有显著正面影响,但实验执行过程中也面临“依从率低”的问题:即使告知了学生和家长该生存在视力问题,仍有一部分学生不去视光中心进行进一步视力检查和矫正。研究团队通过因果链分析发现,筛查结束后学生未及时到视光中心就诊、视光中心到诊率低,可能是因为家长重视度不够以及学生对戴眼镜存在认知误区。因此进一步探索改善农村学生视力健康寻求行为的方案仍是后续研究的重点。

后续研究可以在经过验证的有效且可持续解决农村学生视力问题的方案基础上优化研究设计,以进一步提高近视学生到视光中心检查的比例,比如针对提高学生就诊率的“教师激励方案”,基于损失规避和框架效应等行为经济学理论的“家长干预方案”,通过相关研究来探索更有效、更宜推广的农村学生近视问题改善方案,帮助更多有视力问题的农村学生看得更清、学得更好。

### 3. 小结

健康人力资本是人力资本的重要组成部分,学生的营养健康问题需要引起关注。上述研究表明,缺铁性贫血、近视等营养健康问题的解决,对学生的学业表现、心理健康等都有着重要影响,因此也会对未来的人力资本积累产生影响。

在学生缺铁性贫血问题方面,在各类干预方案中,补充含铁的多维元素片是最有效且最简单的干预方案,它不仅能显著改善学生的贫血状况,而且能够提升学生的学业表现,甚至是改善学生的心理健康水平。同时,补贴和激励类的干预,无论是对家长的激励还是对校长的激励,都可以产生显著的正向影响效果。

在学生视力问题方面,在农村地区成立县级政府主导的农村学生视力保护模式(即视光中心)是从根本上解决农村学生视力健康问题的有效方案,可为国家提供可参考、可操作、宜推广的执行方案,为促进儿童青少年身心健康全面发展提供决策依据。

除了研究者在持续探索解决农村学生营养健康问题的有效方案外,相关政府部门也出台、实施了一系列政策措施:《“健康中国2030”规划纲要》中多次强调要重视儿童的营养健康问题,这也是当前精准扶贫的重要内容;为改善学生营养,教育部实施了“农村义务教育学生营养改善计划”;国家开展了关注学龄前儿童的营养健康行动,针对农村0—3岁婴幼儿的贫血问题免费发放辅食营养包。而为改善学生的视力健康,2018年教育部、国家卫生健康委员会等八部门共同制定了《综合防控儿童青少年近视实施方案》,指出要全社会行动起来,让每个孩子都有一双明亮的眼睛和光明的未来(教育部, 2018b)。我们有理由相信,随机干预实验在农村学生营养健康领域的应用能够为教育政策的制定提供更多的科学依据。

#### (四) 随机干预实验在“儿童早期发展”领域的应用

##### 1. 研究问题

人力资本的形成和发展是一个长期积累的过程,把握不同年龄段儿童大脑发育规律及特点对于有效提高人力资本素质有重要意义。根据《柳叶刀》上发表的有关发展中国家儿童发展的系列报告,个体生命最初的1000天经历大脑的快速发育、突触修剪和髓鞘形成等过程,是大脑可塑性最强的时期,也是多数能力(听觉、视觉、语言、认知等)的最佳发展期(Grantham-McGregor et al., 2007)。已有研究指出,遗传因素决定着儿童发展潜能,而营养和有效刺激儿童发展的家庭及社会环境等要素则决定儿童自身发展潜能的实现程度(Walker et al., 2007; Black et al., 2013)。如果在这个关键期缺乏营养和有效刺激儿童发展的家庭及社会环境等要素,将对儿童未来的学业表现、人力资本的形成和积累以及成年后的收入水平产生很大的负面影响(Campbell et al., 2001; Engle et al., 2007; Horton, 2008; Currie & Almond, 2011)。

投资儿童早期发展等预防性的干预措施,远比后期投资学校教育和成人教育等补救性干预措施的成本收益高,同时也更加有效。诺贝尔经济学奖获得者Heckman(2000)的研究指出,平均而言,人力资本投资回报率随着年龄增加会逐步下降,0—3岁婴幼儿早期人力资本投资的回报率最高。因而,在儿童早期大脑发育这一关键期开展干预,对提高人力资本质量至关重要。

许多国家开展了不同形式和内容的儿童早期发展干预项目,并对干预效果进行了科学评估。最著名的有美国的“佩里学前教育项目”和北卡罗来纳州“启蒙项目”、牙买加的“儿童早期发展项目”、哥伦比亚的“以家庭为基础的入户项目”和乌干达利拉社区的“儿童早期干预项目”等。评估结果表明,这些早期发展干预对儿童的语言、阅读和数学等认知能力发展,以及社会情感和课堂行为等非认知能力发展均有显著的正向影响(Reynolds et al., 2010; Walker et al., 2011; Gertler et al., 2013)。从这些儿童早期发展干预项目的长期追踪评估结果来看,干预组样本往往具有更高的受教育水平、更少的特殊教育需求、薪酬更高、住房拥有率更高、接受福利救济的人数更少、犯罪率更低等特点(Boak, 1999; Rivera

et al., 2004; Akman et al., 2004; Gray & McCormick, 2005; Landry et al., 2006; Heckman et al., 2013)。

国内研究者也开展了一系列关于儿童早期发展的研究,并尤其关注中国农村贫困地区0—3岁婴幼儿的早期发展。贫血是制约儿童早期发展的风险因素,有研究指出,我国农村贫困地区0—3岁婴幼儿中有一些存在贫血问题,并主要由喂养行为不当引起(Luo et al., 2014; Zhou et al., 2016; Luo et al., 2017a)。也有研究指出我国农村贫困地区儿童存在早期发展滞后风险的比例较高,这与照料人缺乏科学的养育知识和技能进而缺少有效的亲子互动密切相关,如不给孩子讲故事、不给孩子唱歌、不与孩子玩耍等(Luo et al., 2017a; Yue et al., 2017, 2019; 李英等, 2019; 岳爱等, 2019)。

我国已经充分认识到促进儿童早期发展的重要性。习近平总书记指出,让贫困地区的孩子们接受良好教育是扶贫的重要任务,也是阻断贫困代际传递的重要途径;要对农村贫困家庭幼儿特别是留守儿童给予特殊关爱,探索建立贫困地区学前教育公共服务体系(习近平, 2015b)。我国也相继出台了一系列全国性规划与指导意见以促进儿童早期发展服务更加完善。2017年,国务院印发《国家人口发展规划(2016—2030)》,指出建立完善包括“幼儿养育”在内的家庭发展政策,尽快实施促进儿童早期发展的各项政策措施,探索适合国情的儿童早期发展指导模式,抓好提高人口素质的奠基工程。2018年12月举行的中央经济工作会议明确将“增加对学前教育、农村贫困地区儿童早期发展、职业教育等的投入”确定为2019年度重点工作任务。2019年5月9日发布的《国务院办公厅关于促进3岁以下婴幼儿照护服务发展的指导意见》指出要“加大对农村和贫困地区婴幼儿照护服务的支持,推广婴幼儿早期发展项目”。

农村儿童是我国未来社会劳动生产力的重要力量,政府提供普惠服务是农村贫困家庭接受婴幼儿照护服务与科学育儿指导的重要途径,但如何将全国性政策落地实施,则需要实证研究的基础上提出干预有效、成本可控的可操作方案。国内有研究团队自2012年开始,基于儿童早期发展理论和中国农村贫困地区儿童发展的现状,使用随机干预实验方法,联合儿童早期发展领域专家设计、实施、评估了五个儿童早期发展项目,旨在结合当地实际情况以及农村社区支持环境,探索可推广、能复制、易落地的儿童早期发展服务方案,为政府在农村贫困地区提供普惠性科学育儿指导服务提供参考和借鉴,以进一步推进国家“教育精准扶贫”战略的实施。

## 2. 农村“儿童早期发展”领域的随机干预实验研究

首先,研究团队从制约儿童发展潜能实现的“营养”元素方面开始探索,设计了“营养包补充”项目,为农村贫困地区儿童每天补充一包富含铁及微量元素的营养包并利用随机干预实验评估影响。结果表明,提供营养包干预在短期可以降低贫血、促进儿童早期发展,但长期无显著效果。

其次,研究团队从制约儿童发展潜能实现的“养育”元素出发,设计了“养育师入户开展亲子指导”随机干预实验研究项目,干预内容为经过研究团队培训的养育师每周进行一次入户家访,指导照料人开展科学的亲子活动,评估结果显示该干预可以显著促进儿童早期发展。随后国际救助儿童会在河北和云南两省使用研究团队开发的干预材料,验证了两周一次入户干预同样可以有效促进儿童早期发展。

再次,为使经过验证的有效干预方式能覆盖更多的农村儿童,研究团队在儿童居住相对聚集的农村社区建立了50个村级儿童早期发展活动中心。这不仅扩大了儿童早期干预项目的覆盖面,也为照料人(主要为妈妈或奶奶)提供了一个交流和获得更多社会支持的场所。

最后,对于农村贫困家庭而言,政府提供普惠服务是其接受科学育儿指导的重要途径。为探索由政府主导、科研团队提供技术支持的儿童早期发展服务提供模式是否可行,研究团队联合陕西省宁陕县政府和公益机构在宁陕县建立了20个儿童早期发展活动中心和6个儿童早期发展活动服务点,从而实现了整县覆盖。

研究团队开展的上述随机干预实验研究均在农村贫困地区实施,采用组群层面抽样,随机分配样本村为干预组或对照组。以下为五个项目的具体介绍(见表6)。



表6 “农村儿童早期发展领域”的随机干预实验研究汇总表

序号	时间	地点	干预方式及样本量	干预效果	参考文献
1	201304-201410	陕西	样本:共11个县174个乡镇351个村6—11月龄1802个婴幼儿 对照组:117个村610名婴幼儿 干预组1(营养包组):117个村600名婴幼儿 干预内容:每天补充一包适合6—36月龄婴幼儿的富含铁等微量元素的营养包 干预组2(营养包+短信组):117个村592名婴幼儿 干预内容:在干预组1的干预内容基础上增加信息干预,发送一条提醒每天服用营养包的信息 干预持续时间:干预持续18个月,每3个月送一批营养包,每半年开展一次追踪评估	干预组1在干预6个月后,12—17月龄婴幼儿血红蛋白水平显著提高了1.77 g/L,认知得分显著提高2.23分(0.13个标准差),但干预效果没有持续;干预组2在加入了信息干预后,提高了营养包的依从率,但是相对于营养包干预组,贫血率并没有显著降低	Zhou et al. (2016); Luo et al., (2017b)
2	201411-201505	陕西	样本:共131个村的592名18—30月龄的婴幼儿 对照组:66个村380名婴幼儿 干预组:65个村212名婴幼儿 干预内容:计生专干开展1对1的入户家访活动,指导照顾人和婴幼儿开展有利于婴幼儿发展的亲子游戏活动 干预持续时间和频率:项目持续6个月,每周入户一次,每次40分钟	一周一次入户干预在实施6个月后,婴幼儿认知得分显著提高了0.27个标准差	Sylvia et al., (2018)
3	201509-201609	云南和河北	样本:共43个村的449名6—18月龄的婴幼儿 对照组:22个村227名婴幼儿 干预组:21个村222名婴幼儿 干预内容:社区的健康工作者开展两周一次的入户家访活动,指导照顾人和婴幼儿开展有利于促进婴幼儿发展的亲子游戏活动,传递健康知识 干预持续时间和频率:项目持续1年,每两周入户一次,每次40分钟	两周一次入户干预将婴幼儿认知得分显著提高了0.24个标准差	Luo, et al. (2019b)
4	201511-201906	陕西	样本:共100个村1664个6—24月龄的婴幼儿 对照组:50个村803名婴幼儿 干预组:50个村861名婴幼儿 干预内容:在每个村建立了一个儿童早期发展活动中心,免费为本村的6—36月龄婴幼儿提供一对一的亲子游戏和亲子阅读等活动,3岁以下的婴幼儿可以到儿童早期发展活动中心玩玩具和看书	干预一年后,儿童的认知水平显著提高,两年后,干预显著地促进了儿童的认知、语言和运动发展	
5	201806-201904	陕西	样本:共69个村/社区995名6—36月龄婴幼儿 对照组:499名婴幼儿 干预组:496名婴幼儿 干预内容:在人口密集地区(服务中心所在地1公里交通距离范围内常住6—36月龄婴幼儿达到10名及以上)采用了养育中心与走访相结合的模式,在人口分散地区则采用家访与家庭小组活动相结合的方式,为照顾人和婴幼儿提供一对一亲子游戏和亲子阅读,以及集体故事会和集体游戏活动	干预10个月后,干预组婴幼儿比对照组婴幼儿在沟通能力、精细运动能力和综合发展水平上有显著提高	

研究一：“营养包补充”随机干预实验研究

研究简介

针对农村贫困地区儿童早期发展存在的贫血和认知潜能未充分发挥的问题,国内外已有一些针对

中国农村贫困地区 0—3 岁婴幼儿的随机干预实验,研究发现营养干预可以有效促进农村贫困地区婴幼儿身高、体重等方面的健康发展(Shi et al., 2010; 丁小婷等, 2016; 徐娇等, 2017),而且一定程度上改善了儿童的认知发展水平(Sheng et al., 2019)。基于国内外研究经验,研究团队为了了解西部农村贫困地区 0—3 岁婴幼儿健康、认知等的发展现状,并找到可能的解决方案,于 2013 年在秦巴山区针对 11 个县 174 个乡镇 351 个村的 6—12 月龄婴幼儿开展了为期 18 个月的“营养包补充”随机干预实验研究。

2013 年 4 月,研究人员对样本地区所有 6—12 月龄儿童进行基线调研后,将 117 个样本村随机分配为“对照组”(不接受任何形式的干预),另外 117 个村随机分配为“干预组 1: 营养包组”(接受免费的营养包干预),剩下的 117 个村随机分配为“干预组 2: 营养包+短信组”(不仅接受免费的营养包干预,每天早上还会接收一条提醒家长给婴幼儿服用营养包的短信干预)。两个干预组样本家庭均接受了贫血的原因、后果以及如何给婴幼儿喂食营养包等相关知识的辅导。研究团队使用的营养包是世界卫生组织推荐的为 6—24 月龄儿童治疗缺铁性贫血的营养补充物,含有铁、锌、维生素 A 等多种微量元素(WHO, 2011)。每隔 6 个月,研究团队对样本儿童进行一次跟踪调查,使用贝利婴幼儿发展量表第一版(BSID-I)和家长问卷跟踪测试样本儿童的认知、运动和饮食等情况。研究团队持续跟踪样本儿童至 24—30 月龄。

### 研究结果

研究结果表明,每天给婴幼儿提供营养包,可以在短期内有效降低婴幼儿贫血率,并提高他们的认知发展能力。干预 6 个月后的结果显示,该干预有效改善了 12—18 月龄婴幼儿的贫血状况(提高血红蛋白水平 1.77 g/L)和认知能力(提高认知得分 0.13 个标准差),但是在后续 2 期(每 6 个月为一期)的追踪调查中发现,长期来看该干预对婴幼儿的贫血状况和认知能力没有显著影响,且这一结果与服用营养包的依从率、项目的统计功效以及喂养行为没有关系(Luo et al., 2017b)。

### 研究局限

进一步分析研究结果发现,虽然“营养包补充”干预在短期内有效改善了农村婴幼儿的贫血状况,提高了他们的认知发展水平,但仍面临一个挑战,即依从率低:那些本应接受营养包干预的很多家庭实际上没有给儿童喂食营养包,或者不能坚持每天给儿童喂食营养包。提高照料人给儿童喂食营养包的比例是研究团队需要进一步解决的问题。研究团队也尝试了提高依从率的一些方法,发现给部分提供了营养包的家庭同时提供短信提醒服务时,依从率有所提高(Zhou et al., 2016)。

该研究面临的另一挑战是:干预方式对干预组儿童没有长期影响。此外,虽然在自然条件下,未接受任何干预的对照组儿童贫血率也随着儿童年龄增长而持续降低,但仍有 23.3% 的儿童存在贫血问题。研究团队认为后续研究应关注如何提升营养干预项目的长期效果,以探索进一步降低贫血率的有效方案。

## 研究二:“养育师一周一次入户开展亲子指导”随机干预实验研究

### 研究简介

2013 年底,研究团队在分析了“营养包补充”随机干预实验研究的初步结果后,针对营养包干预实验存在的局限和挑战,结合牙买加国际项目的经验,在与儿童早期发展专家讨论后,开展了营养和养育结合的干预研究。2014 年,研究团队和国家卫生计生委合作,在秦巴山区 131 个村开展了“养育未来:通过抚育干预促进儿童早期发展”随机干预实验研究,将随机选择的 131 个村中的 65 个村随机分配为“入户干预组”,其余 66 个村为“对照组”(不进行任何干预)。在干预组,研究团队开发了一套适合农村 6—36 月龄婴幼儿的亲子活动课程,通过培训使当地乡镇计生技术服务人员掌握亲子活动课程内容,并由这些计生服务人员负责将研究团队开发的亲子活动课程及与其配套的玩具和绘本教具带到农户家里,与婴幼儿及其家长开展每周一次的“一对一”入户亲子指导活动,干预从 2014 年 11 月开始实施,持续开展了 6 个月。

相较于国际上其他组织开发的儿童早期发展干预方案(如牙买加、世界卫生组织和联合国儿基会等),研究团队开发的亲子活动指导课程是适合中国国情、具有中国特色的,内容更加丰富和细化。课程针对6—36个月不同月龄婴幼儿的认知、语言、运动和社交情绪四方面能力的发展,且配套了能够促进粗大动作/精细动作发展、认知能力发展的玩教具,和促进不同月龄婴幼儿认知、语言、社交情绪能力发展的绘本图书包。这套亲子活动指导课程已于2015年进行了著作权登记,并于2017年出版,2019年被中国家庭教育协会推荐为“百部家庭教育指导读物”。该课程中的指导手册加上配套的玩教具和绘本包,是一套有助于中国儿童早期发展的干预材料,未来可配合政府发展战略服务于“一带一路”国家和其他发展中国家。

### 研究结果

评估结果表明,项目有效改善了干预组照料人的养育行为,干预组照料人给婴幼儿阅读、唱歌和用玩具与儿童玩耍的次数显著提高,更重要的是,一周一次的入户干预对18—30月龄婴幼儿的认知发展有更大的促进作用,其认知得分提高了0.27个标准差(Sylvia et al., 2018)。这一结果与世界卫生组织在安徽开展的随机干预实验的结果一致(Jin et al., 2007)。中国发展研究基金会2014年在甘肃省开展了入户家访的随机干预实验研究,研究结果表明,营养与家访相结合的综合干预对贫困地区0—3岁婴幼儿的动作语言等能力发育、体格生长发育、营养状况都产生了显著的改善效果(中国发展研究基金会, 2017)。这在一定程度上证明了,在人力成本不断上升的今天,“养育师入户指导”是一个促进儿童早期发展的可行方案。

### 研究局限

异质性分析表明,当奶奶是主要照料人时,干预对婴幼儿早期发展没有产生统计上的显著影响(Sylvia et al., 2018)。随着城镇化的发展,很多妈妈会选择在孩子18个月后进城务工,因此,如何改善留守在家照料婴幼儿的奶奶的养育行为是研究团队面临的新挑战(Yue et al., 2020)。另外,乡镇计生技术服务人员进行入户干预依从率不高,平均完成率为60%(项目设计的24次入户活动中,养育师平均完成16.4次)(Sylvia et al., 2018)。通过访谈养育师和农户,研究团队发现有两个原因导致入户率低。第一,样本地区很多农户居住分散,养育师不喜欢去偏远农户家;第二,养育师入户时农户不在家,或者农户对项目不感兴趣、不积极配合。这些原因成为研究团队在推广项目过程中遇到的新挑战,因此,还需进一步探索改善方案。

### 研究三:“养育师两周一次入户开展亲子指导”随机干预实验研究

#### 研究简介

研究团队开展的“养育师一周一次入户”干预显著改善了婴幼儿照料人的养育行为,从而改善了婴幼儿的认知能力,但这一方案成本较高。是否有在保证干预效果的基础上缩减实施成本的有效方案?为此,国际救助儿童会联合国家卫生计生委在云南省和河北省的43个村开展了“养育师两周一次入户指导”随机干预实验研究,研究团队经过随机分配,对干预组的449名6—18月龄儿童及其家庭开展了两周一次的入户亲子活动指导。该研究使用的亲子活动课程等干预材料与上述“养育师一周一次入户开展指导”干预研究用材料完全相同,不同之处在于入户间隔时间为两周。

#### 研究结果

研究结果显示,该干预对婴幼儿的认知发展水平有显著的积极影响,他们的认知得分提高了0.24个标准差,但对于运动和社交情绪没有显著的作用(Luo et al., 2019b)。需要强调的是,研究团队还发现,除了乡镇计生干部,当地村级层面的卫生工作人员也可以承担“养育师”的职责,并能够有效开展入户干预项目。

#### 研究局限

与一周一次的养育师入户干预研究类似,该研究同样存在依从率较低的问题。同时,研究团队还发现,不论是一周一次的养育师入户指导还是两周一次的养育师入户指导,干预成本都不低,平均达每



月 300—400 元/家庭(王蕾等, 2019)。因此,探索能够提高干预依从率和降低干预成本的有效方案仍是研究团队下一步努力的方向。

#### 研究四:“村级儿童早期发展活动中心”随机干预实验研究

##### 研究简介

研究团队在测算“一周一次养育师入户干预”研究和“两周一次养育师入户干预”研究的成本效益后,开始探索降低干预成本、进一步扩大干预覆盖面的方案,尝试在农村贫困地区建立村级儿童早期发展活动中心(简称“养育中心”)。自 2015 年始,研究团队在秦巴山区的 100 个村随机选择 50 个村为“干预组”(每个村建立一个儿童早期发展活动中心),其余 50 个村作为“对照组”(不做任何干预)。对干预组的 50 个村级养育中心,卫生部门提供 2 名乡镇计生专干作为养育师,养育师每周三到养育中心指导照养人和 6—36 月龄婴幼儿开展一对一的亲子活动,亲子活动课程与入户干预相同。研究团队在养育中心所在村雇 1 名村民作为管理员,负责养育中心亲子活动的开展和日常管理,具体为周一到周六负责中心开关门、清洁消毒等养育中心管理工作,并于每周一和家长及婴幼儿开展集体亲子活动,每周五和家长及婴幼儿开展亲子绘本阅读活动。此外,照养人与婴幼儿都可以在养育中心的开放时间阅读绘本、玩耍和交流。研究团队在环境创设方面对儿童早期发展活动中心进行“动静分区”的布置,以保障中心环境适合于开展不同类型的干预课程。

##### 研究结果

初步分析结果表明,“村级儿童早期发展活动中心干预”显著改善了照养人的养育行为(包括亲子阅读、唱儿歌、玩耍等),并且对儿童的认知、语言和运动能力(包括祖辈为主要照养人的儿童的各项能力)均有显著的积极影响。

##### 研究局限

虽然初步分析结果显示“村级儿童早期发展活动中心干预”取得了积极影响,但研究团队发现执行过程仍然面临一项挑战:儿童早期发展活动中心的总体覆盖率(覆盖率为每月至少到养育中心签到一次的家庭占总样本的比例)只有 60%。研究团队分析,覆盖率低的原因主要有:(1)距离儿童早期发展活动中心较远的农户不方便到中心参加活动;(2)有些家长会因为一些其他原因(比如做农活或者家务)而选择不参加儿童早期发展活动中心的活动。为提高因上述两个原因而选择不参与养育中心活动的家庭的参与度,结合前期研究经验,研究团队希望政府部门在推广儿童早期发展普惠服务时,可以将“养育师入户干预”和“村级儿童早期发展活动中心干预”结合起来,即在农村贫困地区建立村级儿童早期发展活动中心的同时,对于距离养育中心远的家庭或因其他原因不方便来中心接受亲子活动指导的家庭进行养育师入户指导。当然,这种入户干预与中心干预相结合的模式是否有效,还需进一步验证。

#### 研究五:“宁陕养育未来整县覆盖”干预模式探索研究

##### 研究简介

对于农村贫困家庭而言,由政府提供普惠服务是其接受科学育儿指导的重要途径,但目前尚无落地实施的具体方案。在前期项目基础上,2018 年研究团队联合政府和公益机构实施了“儿童早期发展整县覆盖项目试点”探索研究,旨在识别在农村贫困地区开展婴幼儿早期发展服务的有效方案,探索政府主导实施的、地方全面覆盖的儿童早期发展有效干预模式,这对理清地方政府阻断贫困代际传递、提高我国儿童早期发展水平和未来人力资本发展水平的长效影响机制有重要意义。

研究团队选择了陕西省宁陕县作为试点地区,探索通过政府属地管理(管理和督导体系)免费为全县所有 6—36 月龄婴幼儿及家庭提供服务的新模式。宁陕县委、县政府于 2018 年初成立了宁陕县儿童早期发展工作领导小组,并在教育体育和科技局下设项目管理中心,进行社区动员以及日常管理,为项目实施提供保障,推动项目有序高效运行。

宁陕县地处秦岭中段南麓腹地,地广人稀、高山低谷地貌为主的特点为覆盖全县所有适龄儿童带

来了挑战。为实现全覆盖,结合宁陕人文地理实际情况,研究团队因地制宜,在人口较密集地区(服务中心所在地1公里交通距离范围内常住6—36月龄婴幼儿达到10名及以上)采用了养育中心与养育师入户指导相结合的模式,在人口分散地区则采用养育师入户指导与家庭小组活动相结合的方式。

研究团队持续监测项目实践数据,发现项目的覆盖率和参与率较前期项目显著提升。截止到2019年9月,宁陕县已经运营了20个养育中心和6个养育服务点,培训一线项目实践人员56名(包含52名养育师和4名管理中心干事);累计注册儿童1498人,注册家长2678人;累计签到48226人次,开展一对一亲子课程22077节,开展集体活动3270场次;目前在册儿童1073人,在册家长1978人。2019年3月至9月平均覆盖率(每月至少在养育中心或服务点签到一次)为94%,月平均1次上课率为89%(养育未来一对一课程),月平均2次上课率为75%。

### 研究结果

从2018年7月开始,研究团队通过贝利婴幼儿发展量表第三版(BSID-III)和儿童发育筛查量表(ASQ-III)测评宁陕县所有6—36月龄样本儿童在干预开始前的各项发展水平,在2019年4月进行追踪调研。项目评估初步结果显示,干预开展10个月后,干预组婴幼儿比对照组婴幼儿在沟通能力、精细运动能力和综合发展水平上有显著提高。

### 研究局限

该研究已实现了全覆盖的属地管理且验证有效,目前正在进行干预实施经验总结和标准化干预方案执行手册的编写。下一步计划在更多农村贫困地区验证整县覆盖模式的外部有效性,以期让更多儿童和家庭都能享受经过科学验证有效的儿童早期发展服务。但该研究仍面临一些挑战:地方政府主导的整县覆盖模式是否可以持续有效运营?是否在推广到其他地区后也同样有效?这些问题还需通过持续研究来回答。

### 3. 小结

从国际经验看,促进贫困地区儿童早期发展、消除贫困的代际传递、提高他们未来的竞争力,是国家发展战略的一个重要选择,也是推动经济长期发展和社会稳定最具公平与效率的公共政策。改革开放以来,我国婴幼儿死亡率、低体重儿童的比例持续下降。儿童的认知、语言、社交情绪和其他一系列与人力资本发展密切相关的能力发展在提高整体人力资本质量中显得尤为重要,但是随着中国迈向高收入国家的行列,中国农村贫困地区儿童的早期发展仍然面临着严峻的挑战。针对这些挑战,国家层面也在持续积极应对。自2012年开始实施的“农村贫困地区婴幼儿营养改善项目”,在改善儿童营养和健康方面取得了显著的成效。除了研究者和政府部门,很多社会力量也在研究机构的技术支持下开展了一系列行动研究实践项目,试图从多方合作的角度探索促进农村儿童早期发展的有效方案,以期为国家探索“教育精准扶贫”的有效路径提供可参考的科学依据和可操作的执行方案。我们有理由相信,在该领域开展更多随机干预实验研究,可以为儿童早期发展政策的制定提供更多科学参考。

## 三、聚焦教育精准扶贫:随机干预实验的中国经验

反贫困是整个人类面临的巨大课题。要实现减贫的伟大目标,仍有大量的理论问题需要理清、大量的现实问题有待解决。本文对中国教育扶贫领域开展的随机干预实验研究进行了梳理,展现了随机干预实验方法在理清理论问题、提供政策依据等方面的巨大作用和潜力。2019年诺贝尔经济学奖授予Abhijit Banerjee, Esther Duflo 和 Michael Kremer 三人,也主要是因为他们应用包括随机干预实验在内的“实验性方案”在减贫方面的巨大贡献。诺贝尔经济学奖的肯定引起了国内外对随机干预实验方法的关注和讨论,既有对其积极作用进行肯定的,也有因其局限性而对其排斥否定的。本文通过梳理随机干预实验法在中国教育扶贫领域的实践经验发现,研究者和政策制定者需要对该方法的作用和局限性有清醒的认识,不应因诺贝尔经济学奖的肯定夸大随机干预实验的作用,当然也不应因其实施过程的复杂性和局限性而排斥使用。

随机干预实验是对政策或干预项目是否达到预期成效进行影响评估的一种方法。相对于其他方法,随机干预实验最主要的优势是能够进行因果分析,能够准确评估出项目的“净”影响效果。相对于其他方法,随机干预实验在进行归因时需要更少的假设,且这些假设更容易成立,其评估结果具有更高的内部有效性(Internal Validity)<sup>⑨</sup>。同样是提供证据,相对于其他方法,随机干预实验能够提供更科学、更准确的证据,也因此,该方法在基于证据(evidence-based)的政策制定中越来越受到重视。

从梳理现有旨在减贫的相关研究来看,一些项目评估通常只进行干预样本实施前后观测指标的对比,而缺少反事实对照组,因此有可能得出错误结论。以通过家长培训改善学生的缺铁性贫血状况为例。假设项目实施前学生平均贫血率为27%,项目实施一年后学生平均贫血率降至15%,下降了12个百分点。由此,可能得出“家长培训对改善学生缺铁性贫血有巨大效果”的结论。但事实上,学生缺铁性贫血状况的改善可能受诸多因素影响,而并非仅由“家长培训”带来。例如,在项目实施过程中,学校可能在同时给学生补充含铁的多维元素片,或者在培训家长的同时教师也参加了相关培训、进而改善了学生在校期间的饮食。学生贫血状况的改善到底是由“家长培训”项目带来的,还是由其他干预措施带来的,理清这一问题有重要的政策意义。基于错误证据或不准确的证据设定的政策可能会带来巨大的社会资源浪费。运用随机干预实验方法评估的结果表明,相对于“家长培训”,补充含铁的多维元素片对改善学生的缺铁性贫血状况效果更好。

从这一过程可以看出,虽然诺贝尔经济学奖肯定了“实验性方案”在减贫领域的应用,但仍需要明确随机干预实验本身并非减贫的直接解决方案。更准确地说,随机干预实验是为减贫寻找、发现出相对更好的解决方案的方法。一项减贫方案的优劣并不取决于是否使用了随机干预实验,不是说使用了随机干预实验方法进行评估的方案就一定好,也不是说没有使用随机干预实验方法的方案就不是好方案。因此,不应夸大随机干预实验的作用,而忽视了准确的问题界定、更合理的干预设计等基本问题。

当然,也不应因随机干预实验本身的复杂性而排斥使用。相对于其他研究方法,随机干预实验的使用相对更复杂,例如,随机干预实验需要有对照组,因此需要更多的样本观测。随机干预实验需要基线调研、评估调研等多次观测,而不只是一次观测。随机干预实验对于项目过程的控制要求也更高,例如需要避免因对照组和干预组因相互交流而带来的污染。更重要的是,随机干预实验不像其他研究方法多以旁观者的身份进行监测,而是要求研究者不仅要观测,还要设计项目、参与到项目实施过程中。比如要评估计算机辅助学习干预对学生学业表现的影响,就需要制作相应软件、去学校安装软件、对教师培训使用软件、对使用软件的过程进行监测、对可能的错误进行纠正等。研究者在看到其复杂性的同时,也应充分肯定其相对于其他方法的优势,不应因畏惧实施的困难而排斥使用随机干预实施方法。

并不是说随意地使用随机干预实验方法就能找到好的教育扶贫方案。总结中国实践可以发现,要通过随机干预实验发现、推动实施有效的教育扶贫政策,需要解决好以下几方面问题:(1)合理、精准的干预设计;(2)深入的数据分析;(3)避免威胁内部有效性的因素;(4)应对好项目规模化的挑战,提高项目的外部有效性(External Validity)<sup>⑩</sup>。中国随机干预实验的实践为解决这些问题提供了重要的经验借鉴。

### (一) 问题精准、干预合理

一项干预方案的优劣,首先取决于问题选择是否精准、干预设计是否合理。

#### 1. 问题导向,精准聚焦,选题要有较强的政策性

选择合适的研究问题是确保教育扶贫项目既“精”又“准”的第一步,而确定研究问题的重要原则之一是政策相关性。研究的问题应当是瞄准教育现实的,应当是教育政策制定者(或其他相关人员)关注的、迫切想要解决的现实问题(例如,学生的心理健康问题),或者是制定政策时迫切需要理清的问



题(例如,确定是否要增加对改善学生缺铁性贫血问题的投入,首先需要理清缺铁性贫血问题到底带来哪些危害,是否影响学生的长期健康或学业表现)。因此,研究者在选题时需要教育政策制定者和项目实际执行者的参与。

问题精准聚焦还应明确研究关心的到底是谁的问题,明确项目最终是为谁服务的,尤其需要区别过程对象与最终对象的不同。例如,在改善儿童早期发展水平的项目中,过程对象可以是社区、父母等,但最终对象是孩子。对家长的培训要解决的问题并非提高家长的养育知识,而是要最终改善孩子的发展水平。项目是否有效应以孩子的发展水平最终是否提高为判断依据。

当前聚焦教育精准扶贫,需要关注影响农村贫困地区教育发展的两个根本问题:一是教师发展面临的挑战,二是学生学习资源、学业表现和身心健康等方面存在的问题。教师作为提高学生学业表现的核心因素之一,在教育中发挥着不可替代的作用。而当前我国农村贫困地区的教师发展仍面临环境和资源的限制,许多问题有待回答:教师能力能否通过“国培计划”得到提升?教师能力的提升是否意味着教学行为的改善?能否通过现代信息技术改善传统的教学方式,解放部分教师工作量、提高教学质量?师范生教育能否为农村贫困地区提供高质量且留得住的师资?而对农村学生来说同样面对许多挑战:营养改善计划是否改善了学生的营养状况(比如贫血发生率)、提高了其学业表现?心理健康问题能否得到准确识别及相应治疗?缺乏课后辅导的学生能否在学校得到相应的免费补习?对他们来说遥不可及的现代信息技术能否更近一点?这些技术是缩小了还是扩大了城乡教育的差距?这些问题仍需通过科学严谨的随机干预实验研究来回答。

## 2. 多方合力,设计干预方案时加强多学科合作

虽然当前各研究学科有着各自明确的研究领域,但现实问题是复杂的,一个现实问题的出现可能是多方面的原因带来的。随机干预实验是一项具体的社会行动,社会行动的落实必然要求项目设计要考虑到社会生活的各个方面。随机干预实验本身是以经济学为主要基础的评估方法,但干预项目的设计必然要求多学科交叉。当今教育研究和教育政策的设计,明显表现出从宏观到微观、从静态到动态、从单一到多元的变化趋势,集合经济学、教育学、心理学和计算机科学等多学科的交叉研究已成为教育研究、教育政策设计的发展趋势。

促进多学科合作是提高干预设计质量的关键因素之一。例如,要解决学生的缺铁性贫血问题,首先需要营养学、医学等相关背景学者的参与,需要明确缺铁性贫血发生的生理原因是什么,有哪些可能的解决办法;如果需要进行信息干预,则需要营养学、医学学者帮助设计信息干预手册,明确要向学生和家长传递哪些关键信息;问卷设计也需要营养学、医学学者的参与,帮助明确测量哪些营养健康指标、如何测量等;如果要通过激励制度设计改善这一问题,还需要教育学、心理学等学科学者的参与。解决现实问题的干预设计并非某个学科的学者可以独立完成的,当前的学科划分方式、以学科为主要领域开展研究的模式是开展随机干预实验的重要挑战之一。以现实问题为中心、而不是以学科为中心,促进多学科合作的交叉研究,能有效提高干预方案设计的精准性、可行性。

## 3. 逐步探索,简单渐进

教育产出受多种因素共同影响,但是用一次尝试多种因素的干预方案来解决所有问题是不可取的。例如,对于学生的缺铁性贫血问题,可能的解决方案包括补充含铁的多维元素片、对学生家长或教师进行营养健康教育、对校长进行激励等。但是否应将这些因素“打包”组合成一项干预进行影响评估呢?如果进行“打包”,该干预有影响效果的可能性更高。但这也面临诸多问题:(1)作用机制不清楚。即使该干预被评估有效果,但仍然不清楚哪种因素起了作用。例如,单独看,对家长的信息干预对改善学生缺铁性贫血问题没有显著影响,这表明从“打包”组合的方案中去除这一因素不会影响整体效果。如果可以精简干预内容但未精简,则会造成社会资源的浪费。(2)推广实施面临巨大挑战。复杂的干预设计会对可获得的资源数量、项目监督等项目实施条件提出更高要求。例如,对于这一“打

包”的干预,是否应该满足了提供多维元素片、家长信息干预等条件才可能实施?如果部分县区没有相应人力对家长进行培训,“打包”干预中的其他部分是否也不应当实施了?因此,干预设计应当是简单的,其判断依据是对最终结果的影响机制是否简单和清晰明了。每个随机干预实验的组别应只检验单个因素的影响效果;如果需要检验其他影响因素,则需要另外的随机干预实验或另外的干预组别。这看似会增加项目成本,但其实会更快地厘清问题、找到有效方案,是更可取的一种方式。

要通过一次随机干预实验就找到有效的干预方式并彻底解决问题,通常也是非常困难的。对复杂问题的认知需要在实践过程中发展、完善,因此对干预的设计也需要不断发展、完善。对于评估有效的方案,可以在总结经验的基础上改进以进一步增强其效果。对于评估没有影响效果的方案,也不应彻底放弃,需要分析项目无效的原因,分析因果链条在哪些环节断裂了,思考是否可以通过调整、改进使其变得有效果。

干预设计的“简单渐进”原则将复杂问题简单化、将大问题化解为小问题,通过一步步的累积逐步解决问题,而不是试图一次性、彻底地解决所有问题。围绕一个待解决的问题,通过多次的项目逐步迭代,一步步设计出更有效的解决方案,是中国随机干预实验实践积累的重要经验之一。

#### 4. 干预设计注重理论指导下的创新性

干预方案的设计强调政策导向性,但并不排斥干预设计对理论的应用。某些随机干预实验可能并不强调干预设计的理论意义。例如,对于政府拟实施的一些政策,无论是否有理论意义、理论意义是什么,既然“不可避免”地要实施,就有必要通过随机实验干预方法评估其可能的影响效果。但对于多数问题,在理论指导下设计干预可能会更快地找到有效的干预方案。理论可能预测了某些结果的产生、产生这些结果可能需要的条件等。通过基于理论的因果链分析,可以提前准备干预所需要的条件、控制关键环节避免干预效果耗散等问题。

为检验某理论而设计的干预可能有更广泛的政策意义。例如,针对学生缺铁性贫血问题设计的两项干预是向学校提供补贴和向校长提供激励。单独看,这是两项具体的干预措施,它们如果被评估有效可以转化为具体的政策,但这并不会给其他干预的设计带来启发。但也可以从更深层的理论角度来理解这两项干预:学生的营养健康问题主要是学校资源不足造成的(没有资源改善这一问题),还是激励不足造成的(学校管理者没有意愿改善这一问题)。这样一来,这一随机干预实验的研究结果不仅可以提出解决学生缺铁性贫血问题的具体政策,还可以对其他类似问题的解决提供重要借鉴。例如对于学生的视力问题、体能问题、发育滞后问题、心理健康问题等,在解决这些问题时就可以考虑应优先从资源不足入手还是应从激励不足入手。

此外,干预设计的创新性也应当被重视。一些社会问题可能早已被社会发现,但长期以来没有得到解决,这不是因为之前没有尝试,更可能是因为已有的尝试、已有的解决办法均没有显著效果。要想取得更好的效果,就需要有新的解决思路。这一方面需要对已有的干预方案、解决办法进行梳理总结,了解是否已经有其他机构/组织在尝试解决这一问题?是从哪些方面来解决这一问题的?解决方法是否有效,为什么有效或者无效?在总结已有经验的基础上,结合理论提出新的解决思路、设计新的解决方案,这样才有可能提出更有效的解决方案。

#### (二) 进行深入的数据分析

开展过随机干预实验,发现了可以解决问题的有效干预方案,并不意味着可以直接将其转化为政策。政策制定时可能需要更多细节,例如,政策制定者不只要知道是否有效,还需要知道细节如何执行、如何监管等。这就需要通过随机干预实验数据进行深入的数据分析来提供。

##### 1. 明确关键环节,发现改进项目的线索

基于随机干预实验提出政策时,不仅需要说明哪种干预有效,还需要说明该干预转化为政策时具体如何执行。一个项目的实施、一项政策的落实是复杂的,在实施前必须要明确执行细节:哪些是决

定项目成败的关键环节、需要重点落实,哪部分群体需要重点关注,哪些是管理项目的重要监控指标等。而这些均可以通过对干预的因果链进行分析、对影响的异质性进行分析等来实现。例如,通过建设儿童早期发展活动中心促进婴幼儿发展项目被证明是有效的,但在具体操作中还应明确一些细节问题:中心的建设面积对干预效果的影响,是否应该鼓励妈妈而不是奶奶带孩子来中心,是应该关注家长来中心的次数还是关注他们来中心的总时长,等等。这些分析将为项目的落实提供一套能确保项目有成效的具体操作流程,而不是模糊的建议。

当干预项目未观测到影响效果时,也需要通过因果链分析明确下一步改进项目的方向。项目未发现影响效果,既可能是项目本身没有效果,但也可能项目本身是有影响的只是某些环节未执行好而使影响效果耗散了。以视力干预项目为例,如果发放免费眼镜后干预组学生有眼镜比例、使用眼镜比例均显著高于控制组,但仍然未发现学业表现有差异,则可能说明该种干预对改善学生学业表现本身是无效的。但如果两组学生有眼镜比例、使用眼镜比例未有显著差异,则说明项目没有效果可能仅是执行问题。进一步分析发现,干预组学生并未更多使用眼镜可能是因为:(1)制作好的免费眼镜可能并未及时下发;(2)学生并不珍惜免费得到的眼镜,所获得的免费眼镜很快被损坏了;(3)该年龄段学生视力下降太快,得到免费眼镜后很快需要更换新眼镜,但学生未更换;(4)问题可能在对照组而不是干预组。对照组学生家长重视视力问题,在得到了《告家长书》后就给孩子配了眼镜,从而使得对照组学生使用眼镜的比例与干预组没有区别。每种原因对改进项目的意义是完全不同的。只有厘清项目为什么会失败,才能明确如何进一步改进项目或者完全放弃项目。因此,不仅应该把随机干预实验方法看作评估项目真实影响效果的工具,还应该将其看作动态地、不断改进项目的工具。

## 2. 运用成本效益分析对比干预项目

对于一个干预项目,不仅应关注其是否有效、影响效果大小,还要关注达到这一影响效果所需的成本。成本效益分析反映了实现每单位影响效果所需成本,即成本/效果(或每单位成本可实现的影响效果,即影响效果/成本)。通过成本效益分析可以更直观地对不同干预项目进行比较,为决策者将有限资源优先分配到哪些领域、优先实施哪些政策提供决策参考(Dhaliwal et al., 2013; Evans & Popova, 2016)。例如,虽然同样可以改善学生的缺铁性贫血情况,但相对于对家长进行培训,直接提供含铁元素的多维元素片影响效果更大、成本更低,因此更具有成本效益优势,更值得优先投入。

此外,在设计干预项目时也应注意通过某些设计来提高其成本效益。免费提供的物品可能存在使用率不高的问题。以视力项目为例,可能有部分学生得到免费眼镜后也不使用,这就造成了资源的浪费。相对于免费发放眼镜,提供眼镜兑换券为学生获得免费眼镜增设了一个小小的障碍:花费时间和精力去县城领取这一免费眼镜。如果学生对使用眼镜的意愿很低,则可能因这一障碍放弃领取免费眼镜,从而避免资源浪费。这一小小的障碍设定显著提高了项目的成本效益(Sylvia et al. 2018)。

## 3. 对项目实践进行理论总结和提升

在设计随机干预实验项目时应注意理论指导;在项目完成后,也应注意对项目实践进行理论总结,尤其是在针对某一问题一系列的随机干预实验开展之后。随机干预实验的结果有明确的政策含义,但如果能进一步进行理论总结,则可能使研究结果对现实具有更强的指导意义。中国教育扶贫领域的随机干预实验实践检验或发展了多方面的理论问题,其中有两方面比较重要。

### (1)关于生命周期不同阶段人力资本投资的社会回报率

中国教育扶贫领域的随机干预实验实践表明,生命周期不同阶段的人力资本投资回报率是随着年龄增长而递减的。当以教育人力资本为结果变量时(例如学生的学业表现或者上大学、上高中的概率),针对高学龄段开展的干预项目通常没有针对低学龄段开展的项目有效。例如,对于大学生、高中生的干预远没有对初中生的有效;而对于初中生的干预,没有对小学生的干预有效。当以健康人力资本为结果变量时,同样也表明越早进行干预越有效。中国的随机干预实验实践表明,James Heckman 关



于生命周期不同阶段的人力资本投资社会回报率的曲线(图7)在中国农村地区仍然成立。这一理论的验证为教育扶贫领域进行政策干预指出了重要的方向。

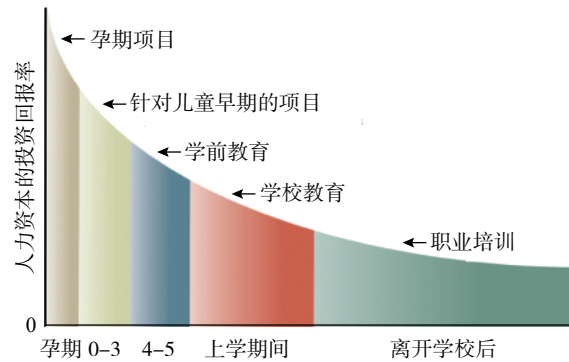


图7 Heckman 曲线: Heckman (2008) 的翻译版本

## (2)关于信息干预的有效性问题的

对于诸多的社会问题,人们最容易想到的干预方式可能是各类信息干预:通过改变干预对象的知识、意识或态度,从而改变其行为。例如,关于垃圾分类,通常的政策干预是对居民进行相关教育、发放宣传页、通过宣传栏进行宣传等。但中国的实践表明,通过信息干预解决问题很难达到预期。比如,关于学生的缺铁性贫血问题,对学生进行课堂教育、给家长发放《告家长书》、对家长进行知识培训、向家长发送短信提醒、对校长进行知识培训等诸多方式均没有显著效果。虽然信息干预是最常用的干预形式,但不应将其作为主要的或唯一的干预形式,而是可以结合其他方法使用,例如只向家长发送短信是没有效果的,但在短信的基础上增加一个有现金激励的测试,就可能有效果。

### (三)避免威胁内部有效性的因素

随机干预实验方法最主要的优势是能够估计出干预项目对结果变量的真实影响效果,这一特点可称之为内部有效性(Internal Validity)。但并非任一随机干预实验均具有内部有效性,内部有效性常常受到威胁,确保随机干预实验的内部有效性需要满足一定条件。中国的随机干预实验实践为处理好内部有效性的威胁提供了重要的经验借鉴。

#### 1. 非预期行为(Unintended Behavioral Effects)

非预期行为是指干预组或对照组样本意识到干预分配结果而产生的非预期反应。这包括两个典型现象。一是霍桑效应(Hawthorne Effect,也称为实验效应),指样本知道自己被选择参加实验而产生个人行为变化,例如更努力工作。二是约翰·亨利效应(John Henry Effect),指对照组样本意识到自己没有分到干预组而更加努力工作。这些都会造成干预项目真实影响效果的偏误(Heckman & Vytalil, 2008)。

解决这些问题可采用单盲甚至双盲的实验设计,即实验参与者不应知道自己被分配组别的情况(单盲)、甚至实验实施者也不知道干预分配情况(双盲)。例如,对于教师绩效工资项目,对于对照组的教师不应让他们提前知道这一干预项目的存在。此外,也可通过类似医学中安慰剂的设计来削弱实验效应。例如,对于通过《告家长书》改善学生的缺铁性贫血问题,在向干预组学生发放《告家长书》的同时,也应向对照组学生发放同样的一份《告家长书》。后者在纸张材质、开头称呼、页数等方面均与前者一致,只是不包含关于缺铁性贫血的关键信息。

#### 2. 不完全依从(Imperfect Compliance)

尽管样本随机分配到干预组或对照组,但会出现被分到干预组或对照组中的样本未完全遵守分配的情况,即依从性问题(Gertler et al., 2013)。这可能是由错误的项目执行造成的,例如在免费眼镜项目中,由于执行中出现偏差,分配到干预组的学生并未收到免费眼镜,而这批眼镜被错误地分发给了对照

组的学生。不完全依从也可能是由非项目原因造成的,例如对于对照组有视力问题的学生,即使未发放免费眼镜,他们也可能自己配眼镜。

为提高项目依从性,研究者可采用提高干预方案的可获得性、改变随机层面等相关实施策略,尽管如此,也仍然难以完全避免这一问题。这种情况下,研究者可通过识别项目依从样本个体与非依从样本个体,通过工具变量方法估计项目的局部平均干预效果。

### 3. 样本污染或溢出效应(Spillovers)

在随机分配后可能出现干预组样本影响到控制组样本或者相反的情况。例如,如果干预组和对照组学生同在一个学校,干预组收到可咀嚼的多维元素片的学生可能会分给控制组的学生吃,或者告诉控制组的学生自己去买这种多维元素片。这种情况的出现会在项目评估时低估项目的影响。

为避免样本污染,一是要改变随机分配的层次,例如不是在学生层面进行随机,而是在学校层面进行随机,从而保证同一学校的学生是被分配在同一干预组别中,从而减少不同组别学生接触的可能性;二是注意抽样方法,例如,在选取学校时,可以每个乡镇只选择一所学校而不是多所学校,由于地理上的隔离,一所干预组学校的学生较少有机会接触到另一个乡镇对照组学校的学生;三是在项目实施过程中减少不同组别接触的机会,例如,如果需要对对照组学校和干预组学校的负责人进行缺铁性知识的培训,应该分开、分批次培训,而不是集中一起培训(即使这样可以节省成本),同时,也应避免将对对照组的联系人和干预组的联系人加入到同一个工作群中。如果实在难以避免这一问题,目前也有一些实验设计的技术性方法至少可以测量出这一效应的大小(Banerjee et al. 2017)。

### 4. 样本流失(Attrition)

在随机干预实验实施过程中可能出现样本中途退出或在评估调研时未能追踪到等情况,即样本流失。这可能带来两方面的问题:一是如果流失的样本与最初样本有显著差异,例如实验开始后,最需要眼镜的、近视度数高的学生因学习成绩差而辍学,则评估时可能低估了项目的影响,因为最需要眼镜的这部分学生的影响效果未包含在内;二是如果流失的样本在干预组和控制组之间有差异,例如,由于获得了免费眼镜,干预组辍学的学生显著少于对照组,此时干预组与对照组不再具有可比性,通过随机干预实验估计出的项目影响效果也将是有偏差的。

为避免样本流失问题,一是在项目开始前应提前估计样本流失的可能性,考虑是否可以提前将流失可能性高的样本排除在样本框之外。例如,对于婴幼儿早期发展的干预,由于农村样本流动性较高,可以在基线调研时询问样本的流动意向,将未来6个月内(干预期为6个月)有外出务工打算的样本暂不纳入样本。二是对样本流失提前做好应对预案。例如,留下家中多名联系人的联系方式,避免样本对象因更换手机号而联系不到等问题。当样本流失问题不可避免地发生以后,可以通过相关分析来评估这一问题可能的影响。研究者可通过分析检验流失样本与非流失样本是否存在差异、流失以后的干预组和控制组是否存在差异等来评估样本流失的影响。

### (四) 应对好项目规模化的挑战,提高项目的外部有效性

在上一节中我们提到,随机干预实验结果的内部有效性受到多重挑战,这些挑战需要在随机干预实验设计、执行和分析过程中进行控制和应对。那么,是不是随机干预实验很好地应对了内部有效性的挑战(也就是说,我们可以将干预组和对照组的结果差异归因于项目干预的影响),其结果就可以在更大范围的人群里进行规模化复制?显而易见,答案是否定的。虽然在理想条件下,通过寻找一个环境和实施伙伴来验证一个新的社会干预方法的概念是有意义的,因为只有这样的环境和实施伙伴,才有可能采纳所有保证成功的必要步骤(例如,进行干预方法的随机分配等),但这样一个小规模实验项目的结果,并不一定能很好地预测大规模实施类似项目的成本和影响,尽管它提供了相关信息(如干预效果及其作用机制等)。事实上,大规模的项目无法复制小型随机干预实验结果的情况并不罕见(Banerjee et al., 2017)。然而,政府的核心目标之一是改善人民的生活水平,尤其是弱势群体的利益,政府需要通过普惠性的公共服务政策来改善这些群体的生存状况。随机干预实验作为一种识别项目效果的

方法需要为政府提供可执行的、可落地的、具有成本效益优势的政策建议,因此随机干预实验的结论需要明确其规模化的影响是什么。

那么,什么是“规模化影响”(Scale-Up Effect)?规模化影响是指项目从小规模研究阶段(通常由项目团队或者非盈利性机构执行)走向规模化阶段(通常是普惠性的并由政府执行)所可能产生的成本效益的变化(AI-Ubaydli et al., 2019; Banerjee et al., 2017)。

项目规模化时,其项目影响与项目成本都可能发生变化。2019年诺贝尔经济学奖得主 Abhijit Banerjee、Esther Duflo 及其他学者总结了小规模项目在规模化时可能面临的六个挑战:溢出效应、市场均衡效应、政治反应、情景相依性、随机化或选择偏误、试点偏差/执行挑战(Banerjee et al., 2017)。这一总结得到了广泛的认可,研究者和政府人员普遍认为它们会影响到项目在规模化实施时的成本与效益。具体来讲,六个挑战及其影响如下(在以下内容中,我们同时也总结应对这些挑战的国内外经验):

### 1. 溢出效应(Spillover Effects)

溢出效应是指干预措施对干预对象的相邻单元的影响。溢出效应既会威胁项目的内部有效性,也会影响项目的外部有效性,从而对规模化产生影响。当干预措施影响的相邻单元是实验对象时(例如对照组样本),溢出效应就会导致内部有效性问题。但当干预措施影响的相邻单元不是实验对象而是更大范围的人群时就会产生外部有效性问题。这种溢出效应可以分为技术溢出效应和信息溢出效应。一些溢出效应与技术有关,例如儿童驱虫项目,如果项目对一个学校的孩子进行了驱虫药的干预,那么临近学校那些没有进行驱虫药干预的孩子也可能受益(Miguel & Kremer, 2004)。一些溢出效应与信息有关,例如实施一项新的技术或干预时(如在线学习、社区孕产妇周末课堂、驱虫蚊帐等),最开始接触它的人群可能不会正确地使用或积极参与,但是当越来越多的人体验过产品或服务以后,他们的朋友和邻居也会经过口耳相传的社会学习而知道该项产品或服务,从而产生强化效应(Reinforcement Effect)。因此,研究者需要识别和厘清溢出效应可能带来的规模化影响。

在经济学中,溢出效应已经被广泛关注,也有不同的方法可以用于测量溢出效应。例如,在“养育未来”整县项目中,项目通过在问卷中设计蓝色染料(Blue-Dye)问题来测量溢出效应,这些问题是只有干预组对象知道的特别信息,通过询问干预组和未被干预的临近单元对于这些信息的掌握程度来测量溢出效应的大小。小规模实验也可以通过实验设计来测量溢出效应。例如,Duflo 和 Saez(2003)运用两层随机过程(Two-Stage Randomization Procedure)的方法在村层面随机分配干预比例,然后再在各村内部随机分配干预组和控制组,以测量退休存款决定的信息溢出效应。但是,并非所有的溢出效应都可以通过小规模实验进行测量,特别是当溢出效应高度非线性(Nonlinearity)时。非线性的溢出效应是指一个干预需要足够多的参与人数才能对临近单元产生影响,例如 Tarozzi 等(2014)进行的驱虫蚊帐项目。因此,当溢出效应高度非线性(Nonlinearities)或存在转折点(Tipping Points)时,研究者需要通过大规模的实验来进行溢出效应的测量。

### 2. 市场均衡效应(Market Equilibrium Effects)

市场均衡效应也可以被理解为是溢出效应的一种。市场均衡效应是指当一项干预措施被大规模实施时,它可能会改变市场的性质。通常来讲,小规模实验在很多情况下与部分均衡分析是一致的:所有市场价格都可以假定保持不变。相比之下,一项大规模的实验——比如全国性的政策干预——可能会影响工资和土地等非贸易商品的价格,而这些价格的变化可能会影响整个项目的净收益(Banerjee et al., 2017)。

Banerjee 等(2017)指出,规模化情况下的项目净收益可能减少也可能增加。例如,奖学金项目通常在小规模实验中能够提高一部分人的教育水平,从而改善他们的收入状况。但是,当项目在人群层面实施时,由于增加了市场上高质量劳动力的供给,供大于求的经济规律会使得工资水平降低,因而在规模化情况下,项目对于收入改善的效果会小于小规模项目的效果(Heckman et al., 1998; Duflo et al., 2017)。再比如,提高收入的项目会产生乘数效应(Multiplier Effect),它不仅会改变被干预者的收入,也



能促进他们的消费,从而影响这些消费品供给方的收入,因此大规模项目中,国民收入的均衡效应要大于小规模实验对收入的影响。

近期的研究采用了不同的方法来应对市场均衡效应的挑战。一种方法是建立模型并进行模拟(Townsend, 2010)。第二种方法则是通过实验设计来测量市场均衡效应。其中比较常见的是通过两层随机过程(Two-Stage Randomization Procedure)在市场层面先随机分配干预比例,然后再在各市场内部进行干预的随机分配,以测量市场均衡效应(Crepon et al., 2013)。另一种实验则是通过在市场层面进行随机并通过市场层面的结果来测量市场均衡效应(Muralidharan et al., 2016)。

规模化情况下的项目的执行成本也可能受到影响,比如前面提到的养育未来项目,如果政府要在农村地区为0—3岁儿童及其家庭提供普惠性的公共服务,那么它需要雇佣大量的一线工作者来执行项目,然而由于目前尚没有成体系的培养体系,此类人员的供给几乎为零,因此该政策将面临招工难的问题。因此,研究者在进行政策建议时应当在成本效益分析中纳入该部分的人力成本。

### 3. 政治反应(Political Reactions)

Banerjee等(2017)提到,随着项目规模的扩大,政治反应可能会有所不同,包括反对或支持项目的反应。有关政治反应的试点项目(比如反腐败)通常比典型的验证干预方法或概念的研究要大得多,因为它涉及的群体要足够大才能产生政治影响。一个小规模试点研究可能会没那么艰难,但这种政治上的反应可能不会显现,因而容易被忽视。然而在其他情况下,试点项目可能比规模化项目更容易受到攻击:因为他们要接受审查,所以很容易被终止。因此,通常需要一定规模的实验才能激发政治反应,并进一步找到解决办法。

在中国的实践中,与政府进行渐进式的合作是一种有效方法。陕西师范大学教育实验经济研究所研究团队及其合作者经过多年的实践总结了研究项目与政府合作的三种渐进模式,即政府观察模式、政府部分参与模式及政府全程参与模式。通过渐近的模式与政府建立信任和合作的关系,有助于逐步达到政府全程参与的效果。

### 4. 情景相依性(Context Dependence)

项目评估通常会在几个精心挑选的地点和特定的组织中进行,其结果可能取决于干预地点或实施组织的一些可观察或不可观察的特征,因此,它们通常不能扩展到不同的地区或组织(即使是在同一个国家)。

解决情景相依性问题的一种方法是对在不同地区或不同组织实施的重复实验进行系统综述。重复实验使研究人员能够理解干预过程中究竟是哪些关键情景因素对直接干预效果产生了作用。例如,Banerjee等(2017)提到的Cochrane Collaboration对健康领域的实验综述,对学龄儿童驱虫、饮水质量以及驱虫蚊帐等项目的重复实验的综述,American Economic Association的注册随机干预实验与数据的公共存档,以及系统元分析(meta-analysis)方法所产生的研究结果。

另一种方法是通过已有实证结果建构理论,利用理论模型的预测指导实验设计,以求验证理论预测,为政策实施提供更为系统有效的指导。例如,Kremer和Glennerster(2011)关注了价格敏感性对预防性保健产品使用的影响。他们通过理论模型构建了不同的理论解释,如流动性约束(Liquidity Constraints)、信息缺乏(Lack of Information)、非货币成本(Nonmonetary Costs)、现时偏差(Present Bias)和有限注意力(Limited Attention)。Dupas和Miguel(2017)则总结了验证这些理论的随机干预实验证据,他们发现一些实验结果支持了Kremer和Glennerster的部分理论,另一些实验结果则证伪了部分理论。

在中国教育领域的随机干预实验的实践中,理论指导下的重复性实验也为政府决策提供了清晰明确的建议。

### 5. 随机化或选址偏差(Randomization or Site-Selection Bias)

Banerjee等(2017)提到,在小规模实验中,同意参与早期实验的组织或个人可能与其他人群不同,Heckman(1992)称之为随机化偏差。

这种担心有三个不同的可能来源。首先,愿意参加随机干预实验的组织(甚至政府内部的个人)通常是例外。第二,当人们选择接受干预时,那些更有可能受益的群体也更容易接受干预,那么随机干预实验的估计效果仅适用于依从者(那些会因干预而产生反应的人),因此这些结果可能不适用于更广泛的人群(Imbens & Angrist, 1994)。第三,选址偏差的产生是因为一个组织为了能够使干预产生有效的结果而选择了一个地点或一个影响特别大的子群体。

通常的解决方法是通过与最终会执行项目的政府或者机构合作,在一个更大的范围内随机抽取具有代表性的人群来进行实验。更大范围和更大规模的实验不仅能够保证抽取的样本更具有代表性,还能帮助研究者在统计上识别出对于不同群体的异质性干预效果。在本文中列举的养育干预项目和营养健康干预项目都是通过这种方式来解决随机化偏差问题的。

## 6. 试点偏差/执行挑战(Piloting Bias/Implementation Challenges)

一个大规模的项目将不可避免地由一个大规模的政府机构来管理,因此,在试点项目中的严密监测和督导在规模化中通常不再可行,或者需要特别的努力。许多研究发现,非政府组织或者研究团队与政府在执行方面存在差异(Banerjee et al., 2007; Barrera-Orsorio & Linden, 2009)。

因此,一个研究项目需要识别项目的关键步骤和核心因素,这些干预的核心部分(或“不可协商的部分”)应该在规模化开展之前进行详细说明,以确保在实施时将执行偏差最小化。那么如何识别出项目的关键步骤和核心因素并减少执行偏误呢?在国际实践(Banerjee et al., 2017; Al-Ubaydli et al., 2019)以及中国实践中,我们总结出以下几点重要经验。

第一,项目设计需要梳理项目从投入到结果的因果链,以及关键的监测和督导时间及其指标。这个过程不仅需要研究团队梳理已有文献,还需要项目相关利益方的共同参与。

第二,需要通过未来会执行政策的政府或机构合作,在更大范围内进行实验,并详实记录执行的过程数据。这是非常关键的内容,然而在现有的很多实验研究中(至少在已发表的研究中)很少有记录执行过程数据的研究。一方面可能是因为记录这些内容的过程繁杂(比如需要一线人员通过手动记录,然后再人工录入),另一方面是因为人为记录的信息可能因为利益相关而不真实,导致收集的信息不会发挥作用。从实践中可以发现,记录方面的问题完全可以通过技术手段进行解决。比如在前面提到的“养育未来”干预项目中,家庭参与儿童早期发展中心活动的记录数据可以由人脸识别系统自动记录和生成,这有效保证了项目干预过程数据的可及性,以及内容的完整性和真实性。

第三,需要通过项目执行手册等使执行方理解项目干预有效的作用机制。只有项目执行者充分理解了成功或失败的关键因素才能更好地减少执行偏误。

第四,项目的研究者应当在项目的实际大规模开展中发挥重要作用,比如起到“脚手架”的作用(即研究团队从深度参与到逐步退出的过程)。研究者应当帮助执行方逐步独立执行项目,掌握执行的核心要素和步骤,并设立研究者逐步退出的机制。

总而言之,研究者应充分认识到项目规模化的挑战,并在项目设计、执行和规模化过程中应对这些挑战。具体来说:情景相依性问题需要通过复制来解决,并且最好是在理论的指导下来完成;市场均衡效应和溢出效应可以通过将这些效应的估计纳入研究设计中,或通过均衡发生时进行大规模实验来解决;随机化和试点偏差可以通过与最终会执行项目的政府或机构合作,在一个足够大的范围内来尝试解决,这需要详细记录成功和失败的过程数据,并帮助执行方掌握执行的关键步骤,也需要详细记录成本数据,并进行规模化的成本效益估计。

## (五)小结

从各国的实践经验来看,成功地设计和完成一项随机干预实验需要特别注意以下两点:

一是要重视对项目作用机制的分析。在随机干预实验的设计、执行及分析等各个环节中,依托因果链进行的机制分析都发挥着重要作用。一个被很好设计和执行的随机干预实验不仅能够厘清项目与潜在结果的因果关系、准确评估项目的影响效果,而且能够打开政策影响的“黑箱”,厘清项目影响

的作用机制。因此,基于因果链设计进行作用机制分析无论是在项目的设计、项目执行过程中,还是在项目评估结束的数据分析时都发挥着重要作用。

在随机干预实验中进行作用机制分析,需注意以下几个方面。从操作流程上看,作用机制的探索体现在随机干预实验的全过程中,而不只是项目结束以后的数据分析中。在项目设计时,就应基于变化理论,理清项目的因果链:每项干预的投入、活动、产出、短期结果、长期结果分别是什么?每个环节的测量指标是什么?各环节成立的假设条件是什么?具体地说,要探索作用机制,在问卷设计时就要明确收集哪些数据、关注哪些指标,在项目执行时要明确重点监测哪些环节,在数据分析时要明确如果项目没有效果,可以检验哪些假设以发现改进项目的线索。

从分析内容来看,作用机制的探索既可以指向实践、服务政策,也可以指向理论、促进创新。在实践层面,可以通过对中间过程变量的变化情况、各环节的假设条件是否成立等进行监测、分析,来明确项目执行过程中的难点,及时调整监督管理的重点。例如,对于没有影响效果的干预项目,要检验哪些应发生改变的中间变量未发生改变、哪些假设条件未如预期一样成立,这对于不断改进项目、形成项目的标准化操作流程、提高项目推广后的有效性具有重要意义。在理论层面,可以根据理论假设设计干预项目(例如,基于理论的影响评估),从多个竞争的理论中检验哪一个理论成立,例如学校的激励问题和资源问题是相互替代还是相互补充;也可以通过多个随机干预实验的对比,总结、提炼出理论,例如,对学生缺铁性贫血问题进行信息干预难以达到预期效果,这可能是由于健康教育的KAP(Knowledge - Attitudes - Practice)理论的某个环节难以实现。

从识别方法上看,随机干预实验可以通过以下几种方式进行作用机制的探索。一是项目的干预设计。例如,在养育项目中,研究人员向干预组提供了特殊信息,这些信息只有干预组对象才知道。因此在评估时向对照组同样询问这些特殊信息时,便可以明确是否存在溢出效应。在探讨解决学生缺铁性贫血的激励问题和资源问题的关系时,随机干预实验可以使用交叉设计,在激励组和补贴组之外增加同时包含这两项干预的干预组,从而可以检验两者之间是相互替代还是相互补充的关系。二是项目的异质性分析。对校长进行缺铁性贫血的信息干预时,对项目的影响效果根据学生是否住校进行异质性分析,则可以明确这种干预效果是通过直接改善学生在校期间的饮食行为实现的,还是通过改善学生和家长的知识和行为实现的。三是对一系列实验的结果进行对比分析。通过信息干预改善学生的贫血状况,涉及接受信息的对象、接受信息的形式、接受信息的频次等多个具体环节,而针对该问题设计的一系列实验恰好瞄准了不同的环节(接受对象从家长到校长,接受频次从一次到两次再到多次,接受形式从宣传页到现场培训再到短信等),这些实验分别检验了信息干预的不同环节,对于理解信息干预的作用机制有重要意义。最后,也可以基于因果链收集中间变量信息,运用中介分析方法探究和对比不同中间变量与最终变量的相关性强度,以识别与项目效果最为相关的因素,从而得出项目效果的产生机制。

二是研究者需要理解、接受和积极应对随机干预实验实践性的特点。一般的研究是思维性的,重在理论构建、数据分析、假设检验等。而随机干预实验不仅是思维性的,更是实践性的。从方法来看,随机干预实验属于行动研究,因其极强的政策导向,随机干预实验也被视为“政策模拟”。随机干预实验的开展不仅需要思维层面的理论思考与分析,更需要实践层面的具体执行与落实,需要去思考和解决各种现实的、有时可能是非常琐碎的问题。某些理论研究者可能不屑于思考和面对一些琐碎的现实问题,例如,如何建立一个工作群联系对照组和干预组的联系人。但这些现实问题能否处理好,会极大地影响随机干预实验的成败。如果把对照组和干预组的联系人放在同一工作群中通知信息,很可能出现严重的样本污染问题。

因此,将一项随机干预实验称为一项复杂的工程并不为过。要通过随机干预实验方法找到解决社会现实问题的有效方法、真正推动社会的进步,需要根据实践的需求,不断进行方法层面的理论创新和实践层面的经验总结,解决好内部有效性、外部有效性等各种问题,做到理性分析、大胆假设、精准



设计、高质量执行。

#### 四、随机干预实验的全球实践对未来教育精准扶贫的启示

当越来越多的国家和个人将资源用于解决发展等民生问题后,也会有越来越多的人会拷问“我们的钱花得是否值得”。这就是干预的有效性问题的思考。这样的思考往往是考虑项目需要在更大范围和规模推广的情况下引出的。

教育精准扶贫,核心是精准。回顾全球在教育精准扶贫方面的工作,一个简单而重要的理念是,没有证据就没有改善。提高减贫能力,需要在政策制定和项目设计中,通过实证证据进行验证。

证据的重要内涵是准确,而随机干预实验是提供准确证据的重要手段。如前文所述,传统的评估方法囿于内生性问题,结果往往存在偏误,借鉴意义受限。随机干预实验作为一个跨领域和跨学科的方法,因其引入外生的随机因素,往往可以得出无偏结果,揭示因果关系,并在此基础上开展公共政策制定所需的成本效益分析。

2019年诺贝尔经济学奖的官方颁奖词是,“表彰其在全球扶贫问题上使用的实验性方法”。以 Abhijit Banerjee、Esther Duflo 和 Michael Kremer 为代表的新一代发展经济学家,将复杂的贫困问题分解为更为具体的问题,应用随机干预实验方法,针对这些具体问题设计了可能改善或者解决的机制与方案。

作为世界上最大的发展中国家,中国从教育、健康等多方面以“摸着石头过河”的方式在扶贫领域精耕细作,扶贫工作取得显著成效。在这样的背景下,在农村贫困地区开展教育、健康等多方面的随机干预实验,不仅更好地厘清了从干预政策到项目结果之间的因果关系,也为更大范围的政策制定和干预推广提供了有效工具和可靠依据。这些基于循证科学论证的有效经验总结和减贫措施,提升了政策制定者对学术研究成果的接受程度,极大地提高了整体减贫能力。这些科学的减贫依据也将为发展中国家乃至全球的减贫工作提供宝贵经验。

本节基于各国政府、国际组织、研究机构的相关实践,进行经验总结,希望发掘相关实践对未来教育精准扶贫的启示。

##### (一) 国际经验的启示

国际上,评估项目的有效性是制定政策的重要环节。影响评估不仅使项目给当地带来实际效应,还能够通过积累知识影响全球发展议程。政府机构如美国国际开发署(United States Agency for International Development)、英国国际开发署(UK Department for International Development),大型基金会如比尔及梅琳达·盖茨基金会(Bill & Melinda Gates Foundation),国际组织如世界银行(World Bank)、泛美开发银行(Inter-America Development Bank)、亚洲发展银行(Asia Development Bank),会资助甚至成立专门的影响评估部门或基金会。美国麻省理工学院的贫困行动实验室(J-PAL)和贫困行动创新组织(IPA)是目前全球最大的两个致力于推动影响评估和实验经济学方法应用的组织。过去十几年,J-PAL 和 IPA 用实验经济学方法开展了大量影响评估研究,为发展中国家发展政策的制定提供了大量的实证依据。

国际评估协会组织(International Initiative for Impact Evaluation, 3ie)成立于2008年,是国际上最有政策影响力的影响评估团队之一。3ie 致力于为发展中国家提供经过科学影响评估验证的政策建议和项目方案。该组织主要通过在国际发展领域资助高水平影响评估项目、提高项目评估质量、生产影响评估项目综述、推广对证据的使用来达成其目标。

3ie 不对项目干预本身进行资助,而只资助影响评估。目前该组织已在超过50个国家和地区资助超过300项影响评估研究,总计1.25亿美元。<sup>⑩</sup>该组织基于这些研究,发表大量影响评估报告、项目评述和政策简报,是利用影响评估促进全球发展的重要平台。目前,3ie的工作主要聚焦于非洲、东亚和拉美地区。在我国的研究则集中于教育和公共健康领域,合作机构主要分布在中西部地区。

在国际组织的推动下,中低收入国家在制定新的教育扶贫政策时很大程度上依赖基于随机干预实验的影响评估结果。以美洲开发银行为例,其26个客户国如果想向美洲开发银行申请贷款推行新的

教育扶贫项目,在立项前就需要以有效性为前提进行项目设计<sup>⑫</sup>。项目设计需要提供合理解决问题的方案,而这些解决方案要能基于实证研究结果论证该项目的潜在有效性。

## (二) 中国实践的启示

### 1. 项目设计需要尽量简单

对于一个社会问题,人们可能想到的干预是多方面的、多层次的,其解决通常需要整合社会资源、上下联动、多方参与。但由于开展教育领域的随机干预实验最终是为了推动教育政策的改善,如果实验方案过于复杂、对实施者的要求过高,则会给后续政策推广造成一定的困难。

例如,对于学生的营养问题,可能的解决办法包括:(1)让学生每天服用一片含铁的维生素片;(2)财政补贴改善学生饮食;(3)通过激励让学校领导更重视学生的营养问题(例如,设置流动红旗);(4)给学生开展健康教育课程;(5)给学生定期进行体检;(6)给家长进行营养健康知识培训。这些干预协调了各方面资源,调动了各方面的积极性,政府、社会(企业)、家长、学生本身都参与进来了。这些干预使用了多种方法,包括激励、物质干预、制度设计等。这些干预可以同时实施,作为一个“组合拳”去解决学生的营养问题。

尽管使用随机干预实验可以把“组合拳”作为一个整体进行影响评估,但当项目的组成办法过多时,其作用机理难以厘清。在随机干预实验中,如果要严格评估每一个办法本身的有效性及其与其他办法的有效性的差异,样本量需要几何程度的增加——往往由于成本和样本数量所限在现实中难以做到。

无法厘清项目机理对项目大规模推广复制的有效性和可行性都提出了巨大挑战。除了需要更多的资源,相比于一个适用于一刀切执行的项目,复杂项目在执行过程中的有效性要大打折扣。“组合拳”中,真正起作用的是什么是全部都有用,还是有些干预其实没有发挥作用,甚至相互冲突、抵消?如果在政策推广中,不能完全复制“组合拳”,只执行其中几项措施,项目还会有效吗?在规模化阶段,大规模执行无效的项目就是一种资源的浪费。

### 2. 政策制定与推广需要成本效益分析的证据

除了通过建立更为直接和严谨的评估标准,如何应用影响评估/随机干预实验研究结果为政策制定者提供更为有效的建议亦十分重要。在教育扶贫政策制定过程中,实现目标的干预方案往往不止一种,在众多方案中进行选择以及更大范围的项目推广都需要有关项目成本效益的研究证据。因此,进行基于影响评估结论的成本效益分析极为重要(White, 2009)。

成本效益分析通过分析比较项目的全部成本和效益来评估项目所产生的价值。将这一分析方法运用于教育扶贫公共政策制定过程中,可以告诉我们每一分钱在教育扶贫项目中产生的价值。不同项目中的对比可以帮助决策者在多种政策或项目中做出选择,以实现在公共政策实施中用最小的成本获得最大的收益。

以视力项目为例。该项目采用了信息干预、直接发放免费眼镜干预和眼镜兑换券干预三种形式,其目标是通过提高学生的戴镜率改善学生学业表现,标准化数学测试成绩的提高是这个项目的最终结果指标。干预的成本包括项目成本、税收成本(假设该项目由政府部门组织实施,成本由财政性税收承担,即税收成本)和家庭成本(眼镜兑换券家庭需自行前往县城兑换眼镜的时间与交通成本)。信息干预未能改善项目的最终目标,即改善学生学业表现,说明信息干预方案不具备成本效益。直接发放免费眼镜干预和眼镜兑换券干预均显著提高了学生学业表现,但对比两种干预方案的成本,可以发现,发放眼镜兑换券方案的成本低于直接发放免费眼镜的成本,相较于直接发放眼镜,发放眼镜券更具有成本效益优势(Sylvia et al., 2018)。

### 3. 教育精准扶贫随机干预实验不只是学术研究,更需政府参与

尽管会受多种因素的共同影响,但如果教育政策的制定是基于高质量的实证研究结果,那么政策制定过程本身就可以促进和保障教育政策的科学性。如果教育政策制定者充分理解随机干预实验在

验证项目效果方面的科学性,知道哪些新政策和新项目确实有助于改善教育精准扶贫政策的实施效果,哪些可能是无效的,那么,他们就可以把资源投向那些有效的政策和项目。

教育精准扶贫随机干预实验的最终目标是制定有效的教育领域的扶贫政策,作为政策制定者的政府部门的参与尤为重要。那么作为随机干预实验研究领域的“非专业人士”,教育政策制定者应该如何更有效地参与到教育领域的随机干预实验当中呢?笔者团队根据大量实证研究,总结出了研究者与政府合作的三种渐进式不断探索的实践模式,分别为:(1)观察模式;(2)部分参与模式;(3)全程参与模式。

观察模式,即政府以一个纯粹的观察者的角色来参与解决教育政策关注的某一方面或多个方面的问题。在这种合作模式下,项目团队主要负责开展项目,但从选题到实验设计等各阶段均需得到政府部门的认可,以此促进政府在科学研究证实问题后,作为主体参与到项目下一阶段的实施中。

当政府官员还没有完全理解研究问题本身或某一干预方案时,通常会采取规避风险的做法,以一种谨慎、缓慢的方式来参与项目实验。面对这样的挑战,实证研究应该先于政策制定者的行动。研究团队需要向政策制定者展示详尽的项目报告,重点介绍为何关注该问题、做了什么干预、结果如何以及下一步如何改进等,以此让政府部门意识,到针对这一特定问题研究团队将开展随机干预实验研究,并希望为政府部门提供科学的决策依据,这样一来,政府在下一阶段参与项目时就会减少很多顾虑。

部分参与模式,即政府部门从项目执行的早期阶段开始参与。在这种合作模式下,研究团队将让政策制定者部分地参与该项目的实施,而在项目构思及设计等比较复杂、零散的前期部分,政府部门还是更多地以观察者的形式参与。

对一些问题,已有经国际研究验证的潜在、可行的解决方案,但尚未在中国进行本土化的尝试和改善,这种情况下项目团队需要与政府部门密切合作,以推动政府全面参与项目的实施。从理论上讲,研究团队已经知道某种干预是起作用的,但在方案实施过程中,研究团队需要与政府合作来回答一些基本问题:这种干预是否在当地的政策环境中可行?在已知多种干预方案都有效的情况下,哪一种在本地政策环境下最有效?等等。如果政府部门实地参与项目实施,并且对研究团队评估干预方案的过程进行观察,那么在验证了干预效果后,该方案后续作为政策试点推广的机会也会大大增加。一般来说,项目实施将会逐渐转化为政府的行动,政府部门可以在自己的管辖区域内大规模推广项目成果。

全程参与模式,即政策制定者在项目的早期构思阶段便参与进来,成为项目团队的一部分,参与项目选题构思、实验设计、方案实施、结果分析及政策推广。对于一些教育发展问题,基于国际成熟经验和国内本土化的试点验证,政府已经接受并认可这些成功的项目干预方案,并将作为主体探索下一步推广方案。经过研究团队与政府部门共同探索的推广方案,政府可以独立总结出更适合自己的管辖区域的有效方案,并向其他区域推广。

#### 4. 教育精准扶贫需要更多随机干预实验

从全球范围来看,21世纪后使用随机干预实验等方法的社会发展领域的影响评估研究得到了迅猛发展,尤其是在教育领域。从2006年全球发展中心(The Center for Global Development, CGD)发表的发展中国家社会发展领域影响评估综述报告可知,截止2004年仅有92项教育领域的科学影响评估项目,而10年后这个数字增长到512项(Shi et al., 2015)。国际影响评估协会2015年对1990至2015年间发表的全球中低收入国家开展的教育领域的干预研究(包括RCTs和准实验方法)做了系统的文献综述(Systematic Reviews),该综述关注的238项研究,大多集中在拉丁美洲和加勒比海、撒哈拉以南非洲和南亚,中国仅有不到30项(Snilstveit et al., 2015)。

当然,我们不是说我国教育领域的随机干预实验研究数量不够多,而是希望强调严谨的科学实验能够给政策制定者提供实证决策依据,因此可以更多地使用。从全球发展中国家开展的教育领域的实验研究的结果来看,通过科学的实验设计、严谨的实验执行以及精确的结果分析,不仅能告诉政策制定者哪些干预有效、哪些干预无效,还能清楚地展示出为什么有些干预有效、有些干预却不起作用,从



而可以帮助政策制定者快速筛选出可能的政策方向。而对地域辽阔的中国,仅在一个地方开展政策干预实验是无法惠及所有的贫困群体的,要想将一项政策推广到其他地区,政策成本和异质性影响都是首要考虑的问题,而这些问题都可以通过随机干预实验来解答。因此,借鉴国际成功经验,我们还需要开展更多的教育领域的随机干预实验研究,以识别精准有效的教育扶贫政策,这可能是未来贫困群体教育研究的一个发展方向。

### 附论: 不开展随机干预实验时,教育精准扶贫如何做?

尽管随机干预实验被认为是影响评估的“黄金准则”,国内外已应用该方法开展了大量教育发展方面的研究,在减贫方面应用前景广阔,但它并不完美,仍存在局限性。除文章中提及的在理论和实操方面的局限外,面对复杂的现实环境和各种各样的资源/条件限制,随机干预实验还并不能适用于所有研究问题。本部分将简要介绍随机干预实验在研究问题上的局限性,并更进一步阐述在随机干预实验不适用或没有条件使用的情况下,如何应用准实验方法开展教育精准扶贫的政策研究。

#### (一) 随机干预实验在研究问题上的局限性

随机干预实验在研究问题上,主要有伦理和逻辑两方面限制(Duflo et al, 2007):

##### 1. 伦理原因

有些问题的干预需要实验者有目的地给干预组提供好处却不给对照组提供可能,这不符合伦理要求。比如,在评估教育对人力资本的重要性时,若通过直接开展随机干预实验评估随机分组后教育水平供给的不同对干预组和对照组样本群体收入水平的影响,我们不能禁止对照组样本接受同等水平的教育。因为这样的随机干预实验是不符合伦理要求的,也是无法开展的。

##### 2. 逻辑原因

有时候在研究者开展基线调查数据收集或者随机分配之前,已经实施了与干预类似的政策或项目方案,从逻辑上看,这种情况是不能开展随机实验的。比如,为改善农村学生营养健康状况,国家于2012年推行“农村义务教育学生营养改善计划”,主要是由中央给予经费支持,提高农村学生在校的营养状况。由于政策已经在各地学校推行,若应用随机干预实验评估营养改善对学生身体健康状况及学业成绩的影响,我们无法创造出没有推行政策的对照组样本群体。因此,我们很难通过随机干预实验方法评估已推广政策的实施效果。

#### (二) 准实验方法在教育精准扶贫中的应用

鉴于随机干预实验自身存在一些局限,加上现实存在的环境资源等限制,并不是所有教育发展问题都适合用随机干预实验来提供解决方案。那么,遇到随机干预实验不适用的情况,如何进行教育精准扶贫政策研究?国内外大量实证研究已经为我们探索出除随机干预实验外的其他验证因果关系的“准实验方法”,这些方法可以用于不适于做随机干预实验的一些领域。

随机干预实验被认为是因果推断的“黄金准则”,最关键的就在于可以通过随机实验构建“反事实”作为对照,以验证干预影响。“准实验方法”其实也一样,可以通过满足特定假设条件来构建“反事实”对照组,通过验证两组的结果变化来识别影响。准实验方法内容较多,假设不同,且不同方法在内部和外部有效性上差别较大,使用场景和方法也千变万化。对此做详细探讨已超出本文题目范围,因此本部分仅对几种常用的“准实验方法”进行简要概述,包括工具变量法、断点回归法、匹配法以及倍差分析法。

##### 1. 工具变量法(Instrumental Variables, IV)

假设我们有一个可观测变量,该变量满足以下两个条件:一,这个变量是外生的,即它与误差项不相关;二,与内生解释变量相关。符合这两个条件,我们就可以称这个变量为解释变量的“工具”,即工具变量。使用工具变量法的核心在于工具外生性(Instrument Exogeneity),这意味着通过工具变量估计的结果变量的变化是无偏的,因为工具变量不与其他影响结果变量的不可观测因素相关。但是工具变量的外生性假设是无法检验的,通常情况下要使用该方法,我们需要借助经济行为或反向思考来维持

相信这一假定。不过需要注意的是,工具变量估计不代表平均干预效果(Average Treatment Effect, ATE),而是估计所谓的局部平均干预效果(Local Average Treatment Effect, LATE)。也就是说,工具变量法估计的是对那些由外生工具所引起的干预组或对照组样本的平均干预效果。

Glewwe 等(2016)研究了佩戴眼镜对学生学业表现的影响。该项目首先分析了给学生发放眼镜与学业表现的因果关系。然而发放眼镜不等于学生佩戴了眼镜,考虑到不完全依从的问题,我们还需要无偏地估计真实佩戴了眼镜的这部分学生的学业表现。由于佩戴眼镜这个变量本身具有潜在的内生性,它不仅受到发放眼镜的影响,还可能与家庭到学校的距离以及家长对学生视力和学习的关注程度有关,而这些因素都可能对学生的学业表现产生影响。为识别佩戴眼镜与学业表现真实的因果关系,研究者引入了一个外生变量即“是否得到免费发放的眼镜”,这个变量既与内生的解释变量相关,又与误差项不相关。在这个项目中,发放眼镜仅通过影响学生佩戴眼镜的概率来影响学业表现,不与其他不可观测变量相关,因此不再有内生性问题,可以作为工具变量来识别佩戴眼镜与学业表现之间的因果关系。使用工具变量法分析结果表明,参与该项目的4—6年级学生佩戴眼镜8—9个月就可以将学业成绩显著提高0.41个标准差。

## 2. 断点回归法(Regression Discontinuity Design, RDD)

在某些情况下,我们需要分析处在一个特定变量(通常称为“游动变量”)临界值两端的结果变量的差异,即干预组和对照组分别位于临界值左右两侧的影响评估设计,这称为断点回归法。使用断点回归法的前提是,在这个“游动变量”的临界值两端,干预的可能性呈现突变或不连续的变化。使用断点回归方法识别因果关系,必须同时满足四个条件。第一,用于区分样本的游动变量必须是连续的,比如年龄、考试成绩、收入等。反之,分类变量(例如性别、就业情况、教育程度等变量)则不能用来区分样本。第二,该游动变量必须存在一个“临界值”,使临界值两边的样本分别参与或不参与干预。例如女童奖学金项目,对于所有奖学金申请者,只有家庭资产不高于1.6万元的女童获得了奖学金,则1.6万元就是游动变量(即家庭资产)的临界值。第三,这个临界值必须只可以用来区分该研究项目。如果1.6万元的资产不仅可以决定女童是否收到奖学金,还决定其家庭的医疗保险等其他社会救济,那么我们就不能通过断点回归法来评估奖学金项目的效果。第四,任何个体都不能精确地将其游动变量值控制在临界值的左右,在临界值周围的个体,无法操纵使其落在它所在的临界值的任一边。这就好像使样本个体被随机分配到临界值的左侧或者右侧,从而模拟了随机干预实验的场景。

Park 等(2015)使用断点回归法对就读重点高中如何影响学生学业表现进行了研究。在该项目中,干预方式为是否就读重点高中。区分干预组和对照组的游动变量为样本学生的中考成绩,而录取分数线则为该游动变量上的临界值:高于录取分数线学生可以进入重点高中(即干预组),低于录取分数线学生则只能在普通高中就读(即对照组)。为降低选择性偏误,研究者将样本分组进一步设定条件,即干预组为在重点高中就读且中考成绩略高于录取分数线学生,而对照组则是在普通高中就读且中考成绩略低于录取分数线学生。假设干预组和对照组学生的其他基本特征相似,其差异只在于是否就读于重点高中,那么分析两组学生在三年后高考成绩中的差异就能识别出干预对学生学业表现的影响。

该项目满足使用断点回归法的四个条件:其一,基线调研时,区别样本特征的游动变量(即学生的中考成绩)是连续的;其二,样本特征存在明显的“临界值”,即录取分数线;其三,落在录取分数线两边学生的其他基本个人特征是非常相似的,其差异只在于是否就读于重点高中;最后,在录取分数线周围的学生,其就读的高中只取决于中考分数,而不能人为操纵。使用断点回归法分析结果表明,就读重点高中比没有就读重点高中的学生高考成绩高出0.387个标准差,同时,就读重点高中可以将大学入学率显著提高27.8个百分点。

## 3. 匹配法(Matching)

为了寻找“反事实”对照,在观察数据里面对于每一个接受干预的参与者来说,我们都希望找到一

个没有得到干预的“双胞胎”,这个“双胞胎”是一个对照组参与者,它与实验干预组的参与者一样在除干预之外的其他控制变量水平上有相同的取值。我们就是要通过利用这些控制变量来“匹配”接受干预的样本和未接受干预的对照样本,经过将干预组与对照组“配对”,比较干预组的一个“双胞胎”和对照组的另一个“双胞胎”之间结果的平均值是如何变化的,这种比较是对实验干预平均影响的一个估计。然而,相比于工具变量法和断点回归法,使用匹配法和下面即将介绍的倍差分析法需满足更严格的假设条件。运用匹配法的一个关键性假设是使评估者能控制大量的可观测控制变量,但仍会有一些同样重要的不可观测变量影响着主要自变量和结果变量。因此,为了通过匹配产生相对无偏的估计,研究者需要控制每一个同时与主要自变量和结果变量相关的重要可观测变量。

Bai等(2017)使用倾向匹配得分法评估了高中减免学费政策对提高我国农村学生高中入学率的影响。在该研究项目中,干预组的样本初中生已经受到政策干预,即事前承诺初中生若能考入高中,则减免该生的高中学费。为评估干预效果,研究者构建了对照组,即没有接受到高中减免学费政策的学生群体。考虑到干预已在一个县全面实施(即干预县),研究者选择了另一个县作为对照组样本框,被纳入的对照县与干预县在以下主要特征变量上相似:(1)与干预组样本县隶属于同一个市,且均属于国家级贫困县;(2)与干预组样本县同处于多山地带;(3)农村居民人均收入水平接近;(4)教育经费相近且均由政府承担;(5)学生初中课程内容、高中入学标准以及学费标准一致。这也就保证了对照组与干预组样本县在地理地形、政府财政支持、居民经济状况和教育体系上较为相似,可构建为干预组的“反事实”对照组。通过倾向匹配得分法分析结果表明,高中减免学费政策能显著提高初中生高中入学率21个百分点,显著降低初中生职业高中入学率11.9个百分点。

#### 4. 倍差分析法(Difference-in-Difference, DID)

在寻找“反事实”对照组的过程中,有些干预(个体层面上的干预)可以通过匹配法找到一个没有得到干预的“双胞胎”,而另一些整体层面上的干预,可以通过倍差分析法比较干预前后干预组平均结果和干预前后对照组平均结果的变化来识别因果。倍差分析法的关键假设为“平行趋势假设”,即如果干预组没有进行实验干预,那么干预组的干预前后变化与对照组的干预前后变化遵循相同的趋势,也就是说对照组的平均结果变化代表未经干预的干预组的平均“反事实”的变化。由于倍差分析法依赖于在两个或两个以上的时间点的干预组和对照组结果变量的变化来识别因果关系,因此不要求两组样本在基线时有相同的特征,但必须同时有干预组和对照组样本在干预前后的结果变量的观测值。用干预组干预前后结果变量的变化减去对照组干预前后结果变量的变化,就可以识别出项目产生的影响。

Liu等人(2010)使用倍差分析法研究了中国农村地区小学合并项目对学生学业表现的影响。一些人认为,将偏远地区规模较小的小学合并到规模较大的中心小学,会对学生的学习表现产生一定的负面影响。那么是否合并小学会导致学习成绩下滑呢?研究者在中国西北农村地区选取了62所小学共2446名小学生参与调研,其中,561名学生来自被关闭的小学(干预组A),820名学生来自合并前的中心小学(干预组B),其余1065名学生来自非合并小学(对照组)。该实验研究假设,如果不存在小学合并的情况,两个干预组的学生与对照组学生学习成绩的变化趋势是相同的。那么,分别研究两个干预组与对照组学生在小学合并前后学习成绩的差异就可以识别出小学合并对学生学业表现的影响。分析结果表明,合并小学并不会对学生们的学习成绩产生显著的负面影响。但是,合并时的年龄与学习成绩的变化有显著关系:年龄较大的学生合并后成绩显著提高了,而年龄较小的学生成绩显著降低了。

#### (三) 小结

回顾上文所述,识别一项教育扶贫政策是否实现了精准扶贫的核心在于评估其影响,即测算这项政策的作用对象在接受政策干预前后的结果变化并准确归因。随机干预实验方法与准实验方法,均可广泛应用于在教育领域开展的精准扶贫类政策研究中,通过相关研究来分析干预措施是否对最终结果产生了影响。尽管随机干预实验有其局限性,但仍是因果推断的“黄金准则”,可以识别出干预或政策产生影响的作用机制,有效模拟政策实施效果。在不适用开展随机干预或资源有限没有条件开展随机



干预实验的情况下,可应用准实验方法进行影响评估政策实验来识别有效的教育类政策(或项目),从而促进贫困群体的发展。

需要强调的是,不管是随机干预实验还是准实验方法,均具有其适用性与局限性。研究者在开展实证研究时,更应注重社会问题本身,而非仅关心验证完美的科学实验方法。在开展教育类影响评估时,研究者需结合自己的研究问题和研究项目的实际可能性,判断是应用随机干预实验还是准实验方法来进行政策评估,从而为政府制定教育扶贫政策提供更加科学的决策依据,以进一步促进消除贫困目标的实现。

致谢:

作者感谢以下项目和机构的支持:

国家自然科学基金重点项目(项目号:71933003);国家自然科学基金青年项目(项目号:71703084, 71703083, 71803107, 71803108);高等学校学科创新引智计划(项目号:B16031);教育部人文社会科学研究青年基金项目(19YJC790080);中央高校基本科研业务费专项资金资助项目(项目号:2017CBY017);中国博士后科学基金面上资助项目(项目号:2019M663619);陕西师范大学中央高校基本科研业务费专项资金项目(20SZYB12)。

国家卫生健康委员会干部培训中心;浙江省湖畔魔豆公益基金会;北京三一公益基金会;北京陈江和公益基金会;澳门同济慈善会北京办事处(Macao Tong Chai Charity Association in Beijing);北京情系远山公益基金会;瑞银慈善基金会(UBS Optimus Foundation);国际影响评估协会(International Initiative for Impact Evaluation, 3ie);福特基金会(Ford Foundation);徐氏家族慈善基金会;戴尔(中国)有限公司(Dell China);广达电脑公司(Quanta Computer);中国儿童少年基金会;TAG家族基金会(TAG Family Foundation);宏碁集团(Acer);农村教育行动计划(Rural Education Action Program, REAP);Asia-Pacific Economic Cooperation(AEPC) Digital Opportunity Center(ADOC) 2.0;李谋伟(Bowei Lee)及其家族;南都公益基金会;中华少年儿童慈善救助基金会;依视路中国(Essilor China);同一视界(OneSight)慈善基金会;中山大学中山眼科医院;ThoughtWorks;洛克菲勒基金会(Rockefeller Foundation);郭氏慈善信托(Kwok Charitable Trust);陆逊梯卡(中国)投资有限公司(Luxottica China);上海煜盐餐饮管理有限公司;世界银行(World Bank);广州市好百年助学慈善基金会;北京德方公益基金会;深圳市爱阅公益基金会;携程旅游网络技术(上海)有限公司;北京观妙公益基金会;广东省唯品会慈善基金会;和美酒店管理(上海)有限公司;上海胤胜资产管理有限公司;上海市慈善基金会。

## 参考文献

- 白云丽,易红梅,张林秀,罗仁福,刘承芳.(2015).农村基层医疗卫生机构视力检查服务提供与利用.《中国卫生政策研究》,8(04),75—80.
- 曾满超,丁小浩,沈华.(2010).初中生课外补习城乡差异分析——基于甘肃、湖南和江苏3省的初中学生课外补习调查.《教育与经济》,(02),9—13.
- 常芳,史耀疆,李凡,岳爱,杨斌,杨鑫.(2013).信息干预对留守儿童身体健康的影响——来自陕西省的随机干预试验.《农业技术经济》,(04),119—127.
- 常芳.(2018).《激励与中国农村教育》.北京:社会科学文献出版社.
- 陈宝生.(2019).全国教育工作会上的讲话.[http://www.moe.gov.cn/jyb\\_xwfb/moe\\_176/201901/t20190129\\_368518.html](http://www.moe.gov.cn/jyb_xwfb/moe_176/201901/t20190129_368518.html).
- 陈斌开,张鹏飞,杨汝岱.(2010).政府教育投入、人力资本投资与中国城乡收入差距.《管理世界》,(1),36—43.
- 丁小婷,张飞,何启强,毛宗福,李锐.(2016).江西省农村贫困地区6—18月龄婴幼儿营养干预效果分析.《现代预防医学》,43(020),3703—3705.
- 东方网.(2005).教育部负责人就“农村寄宿制学校建设工程”答问.<http://news.eastday.com/eastday/news/node37955/node37957/node37979/node79277/userobject1ai1341687.html>
- 范先佐,付卫东.(2011).义务教育教师绩效工资改革:背景、成效、问题与对策——基于对中部4省32县(市)的调查.《华中师范大学学报

(人文社会科学版), 50(06), 128—137.

付卫东, 崔民初. (2011). 义务教育学校教师绩效工资政策分析. *现代教育管理*, 02, 66—70.

付卫东, 桂勇. (2010). 义务教育学校教师绩效工资政策实施效果的实证研究——基于中部四省部分县(区)的调查分析. *上海教育科研*, 10, 25—28.

国家统计局. (2001). *中国统计年鉴 2001*. 北京: 中国统计出版社.

国家统计局. (2010). *中国统计年鉴 2010*. 北京: 中国统计出版社.

国家统计局. (2014). *中国统计年鉴 2014*. 北京: 中国统计出版社.

国务院. (2012). 国务院关于深入推进义务教育均衡发展的意见. 链接: [http://www.gov.cn/jjxgk/pub/govpublic/mrlm/201209/t20120907\\_65532.html](http://www.gov.cn/jjxgk/pub/govpublic/mrlm/201209/t20120907_65532.html).

国务院. (2009). 政府工作报告.

胡耀宗, 严凌燕. (2017). 义务教育教师绩效工资政策执行偏差及其治理——基于沪皖豫三省市教师和校长的抽样调查. *教师教育研究*, 29(05), 14—18.

华文娟, 伍晓艳, 许韶君, 万宇辉, 金菊香, 高国朋, 方云, 裴晨璐, 陶芳标. (2013). 我国中部6省城市中小学生疑似近视患病率与纬度分布. *中国学校卫生*, 34(11), 1299—1301.

黄南, 张二震. (2017). 经济转型的目标、路径与绩效: 理论研究述评. *经济评论*, (02), 152—162.

教育部. (2008). 教育部关于做好义务教育学校教师绩效考核工作的指导意见.

教育部. (2010). 国家中长期教育改革和发展规划纲要(2010—2020年). 2010-07-29, 链接: [http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/info\\_list/201407/jjxgk\\_171904.html](http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/info_list/201407/jjxgk_171904.html).

教育部. (2016). 教育信息化“十三五”规划. 2016-01, 链接: [http://www.moe.edu.cn/srscite/A16/s3342/201702/t20170221\\_296857.html](http://www.moe.edu.cn/srscite/A16/s3342/201702/t20170221_296857.html).

教育部. (2018a). 教育信息化2.0行动计划. 2018-04-18, 链接: [http://www.moe.gov.cn/srscite/A16/s3342/201804/t20180425\\_334188.html](http://www.moe.gov.cn/srscite/A16/s3342/201804/t20180425_334188.html).

教育部. (2018b). 综合防控儿童青少年近视实施方案. 2018-08-30, 链接: [http://www.moe.gov.cn/srscite/A17/moe\\_943/s3285/201808/t20180830\\_346672.html](http://www.moe.gov.cn/srscite/A17/moe_943/s3285/201808/t20180830_346672.html).

李克强. (2018). 全国教育大会上的讲话. [http://www.xinhuanet.com/2018-09/10/c\\_1123408400.html](http://www.xinhuanet.com/2018-09/10/c_1123408400.html).

李玲课题组. (2015). 视觉健康政策调整刻不容缓——关于中国国民视觉健康的调研报告. *光明日报*, (005).

李英, 贾米琪, 郑文廷, 汤蕾, 白钰. (2019). 中国农村贫困地区儿童早期认知发展现状及影响因素研究. *华东师范大学学报(教育科学版)*, 37(3), 17—32.

罗仁福, 张林秀, 刘承芳, 易红梅, 史耀疆. (2015). 提高人力资本质量重点在贫困地区. *中国国情国力*, 4, 64—66.

罗仁福, 张林秀, 王晓兵, 易红梅, 史耀疆, Rozelle, S. (2013). 家长健康信息干预对贫困农村学生贫血的影响. *中国学校卫生*, (02), 103—105.

罗仁福, 张林秀, 赵启然, 易红梅, 史耀疆, Rozelle, S. (2011). 陕西贫困农村寄宿学校小学生贫血情况调查. *中国学校卫生*, (10), 111—112.

聂景春, 庞晓鹏, 周国栋. (2016). 信息不对称对农村学生健康卫生服务需求的影响研究——以眼镜为例. *中国农村卫生事业管理*, (12), 91—95.

庞丽娟, 韩小雨. (2005). 农村中小学布局调整的问题、原因及对策. *教育学报*, 1(4), 90—96.

彭泗清, 李兰, 潘建成, 郝大海, 韩岫岚, 韩践, 王云峰, 董博. (2013). 经济转型与创新: 认识、问题与对策——2013?中国企业家成长与发展专题调查报告. *管理世界*, (09), 15—26.

全国教育大会. (2018). 习近平: 坚持中国特色社会主义教育发展道路, 培养德智体美劳全面发展的社会主义建设者和接班人. [http://www.moe.gov.cn/jyb\\_xwfb/s6052/moe\\_838/201809/t20180910\\_348145.html](http://www.moe.gov.cn/jyb_xwfb/s6052/moe_838/201809/t20180910_348145.html).

史耀疆, 王欢, 罗仁福, 张林秀, 刘承芳, 易红梅, 岳爱, Scott Rozelle. (2013). 营养干预对陕西贫困农村学生身心健康的影响研究. *中国软科学*, (10), 53—63.

孙炜红, 张冲. (2014). 中国人口10年来受教育状况的变动情况. *人口与社会*, (1), 79—80.

汪丁丁. (2010). 人力资本与教育. *IT 经理世界*, (10), 106.

王聪. (2017). 绩效工资制度下义务教育教师管理现状与改革突破——基于北京市主要城区调研的思考. *中国教育学报*, 05, 30—35.

王蕾, 贤悦, 张偲琪, 等. (2019). 中国农村儿童早期发展: 政府投资的效益—成本分析. *华东师范大学学报(教育科学版)*, 37(3), 118—128.

王云峰, 田一. (2012). 北京市义务教育阶段学生语文、数学、英语学科学业水平的城乡差异研究(2006—2009年). *教育科学研究*, (7), 51—55.

习近平. (2015a). 携手消除贫困 促进共同发展——在2015减贫与发展高层论坛的主旨演讲. <http://cpc.people.com.cn/n/2015/1017/c64094-27709112.html>.

习近平. (2015b). 脱贫攻坚战冲锋号已经吹响 全党全国咬定目标苦干实干. 人民网, 2015年11月29日01版, <http://cpc.people.com.cn/n/2015/1129/c64094-27867800.html>.

- 习近平. (2016). 在东西部扶贫协作座谈会上的讲话. <http://cpc.people.com.cn/xuexi/n1/2019/0218/c385474-30759690.html>.
- 谢宇, 李汪洋, 靳永爱. (2019). 普及高中阶段教育的背景、意义和挑战. *北京大学学报(哲学社会科学版)*, 56(4), 122—135.
- 熊向明. (2007). 对当前农村中小学布局调整的反思——河南中原地区农村中小学布局调整调查分析. *教育与经济*, (2), 50—53.
- 徐娇, 霍军生, 孙静, 黄建. (2017). 国内外 6—24 月龄婴幼儿辅食营养包干预研究. *中国食品卫生杂志*, (5), 550—555.
- 杨倩, 谢作栩. (2017). 规模扩张、城乡二元与高等教育机会变化——基于吉林省的实证调查. *教育学术月刊*, (1), 34—40.
- 叶怀凡. (2016). 义务教育教师绩效工资政策的执行偏差与矫正. *中国教育学报*, 04, 31—36.
- 叶敬忠, 潘璐. (2008). 农村小学寄宿制问题及有关政策分析. *中国教育学报*, (2), 1—5.
- 岳爱, 蔡建华, 白钰, 汤蕾, 史耀疆, 罗仁福, 罗斯高. (2019). 中国农村贫困地区 0—3 岁婴幼儿面临的挑战及可能的解决方案. *华东师范大学学报(教育科学版)*, 37(3), 1—16.
- 张林秀. (2013). 随机干预试验——影响评估的前沿方法. *地理科学进展*, 32(6), 843—851.
- 张晔, 关宏宇, 李莉莉. (2018). 西北农村小学教师筛查学生视力不良培训效果分析. *中国学校卫生*, 39(10), 136—138.
- 中国发展研究基金会. (2017). 惠育中国: 甘肃华池县项目终期评估报告.
- 中国教育报. (2016). 调查显示: 电脑辅助学习能显著提高学业表现. [http://www.jyb.cn/basc/sd/201606/t20160622\\_663196.html](http://www.jyb.cn/basc/sd/201606/t20160622_663196.html)
- 中华人民共和国国家卫生健康委员会. (2018). 综合防控儿童青少年近视实施方案. 2018-08-30.
- Adams, J. (2007). Living and learning at school: China's rural boarding school construction program and its implications for primary school education. Working Paper. School of Education, Stanford University, Stanford, CA.
- Akman, M., Cebeci, D., Okur, V., Angin, H., Abali, O., & Akman, A. C. (2004). The effects of iron deficiency on infants' developmental test performance. *Acta Paediatrica*, 93(10), 1391—1396.
- Al-Ubaydli, O., List, J. A., & Suskind, D. (2019). The science of using science: Towards an understanding of the threats to scaling experiments (No. w25848). National Bureau of Economic Research.
- Bai, Y., Mo, D., Zhang, L., Boswell, M., & Rozelle, S. (2016). The impact of integrating ICT with teaching: Evidence from a randomized controlled trial in rural schools in China. *Computers & Education*, 96, 1—14.
- Bai, Y., Tang, B., Wang, B., Mo, D., Zhang, L., Rozelle, S., Auden, E., & Mandell, B. (2018). Impact of online computer assisted learning on education: Evidence from a randomized controlled trial in China. REAP Working Paper.
- Bai, Y., Yi, H., Zhang, L., Shi, Y., Ma, X., Congdon, N., ... & Rozelle, S. (2014). An investigation of vision problems and the vision care system in rural China. *Southeast Asian Journal of Tropical Medicine and Public Health*, 45(6), 1464.
- Bai, Y., Zhang, L., Yi, H., Zheng, L., & Rozelle, S. (2017). The impact of an academic high school tuition relief program on students' matriculation into high schools in rural China. *China Economic Review*, 43, 16—28.
- Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, 122(3), 1235—1264.
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., ... & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, 31(4), 73—102.
- Banerjee, A., Glewwe, P., Powers, S., & Wasserman, M. (2013). Expanding access and increasing student learning in post-primary education in developing countries: A review of the evidence. Abdul Latif Jameel Poverty Action Lab.
- Barrera-Osorio, F., & Linden, L. L. (2009). The use and misuse of computers in education: Evidence from a randomized experiment in Colombia. The World Bank.
- Barrera-Osorio, F., M. Bertrand, L. L. Linden, & F. Perez-Calle. (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia. *American Economic Journal: Applied Economics*, 3, 167—195.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106(2), 407—443.
- Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy*, 1(1), 52—74.
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2), 9—49.
- Becker, G. S. (1992). Human capital and the economy. *Proceedings of the American philosophical society*, 136(1), 85—92.
- Behrman, J. R., Parker, S. W., & Todd, P. E. (2005). Long-term impacts of the oportunidades conditional cash transfer program on rural youth in Mexico. In Discussion Paper Series No.122. Göttingen, Germany: Ibero-America Institute for Economic Research.
- Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., De Onis, M., ... & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, 382(9890), 427—451.
- Boak, F. L. (1999). Parent-child relationship, home learning environment, and school readiness. *School Psychology Review*, 28(3), 413.



- Bobonis, G. J., Miguel, E., & Puri-Sharma, C. (2006). Anemia and school participation. *Journal of Human Resources*, 41(4), 692—721.
- Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, 37(2), 231.
- Chadha, R. K., & Subramanian, A. (2011). The effect of visual impairment on quality of life of children aged 3–16 years. *British Journal of Ophthalmology*, 95(5), 642—645.
- Chang, F., Wang, H., Qu, Y., Zheng, Q., Loyalka, P., Sylvia, S., ... & Rozelle, S. (2020). The impact of pay-for-percentile incentive on low-achieving students in rural China. *Economics of Education Review*, 101954.
- Chen, X., Shi, Y., Mo, D., Chu, J., Loyalka, P., & Rozelle, S. (2013). Impact of a senior high school tuition relief program on poor junior high school students in rural China. *China & World Economy*, 21(3), 80—97.
- Cochrane, S. H., Leslie, J., & O'Hara, D. J. (1982). Parental education and child health: Intracountry evidence. *Health Policy and Education*, 2(3—4), 213—250.
- Congdon, N., Li, L., Zhang, M., Yang, A., & Lam, D. S. C. (2011). Randomized, controlled trial of an educational intervention to promote spectacle use in rural China: The see well to learn well study. *Ophthalmology*, 118(12), 2343—2350.
- Congdon, N., Wang, Y., Song, Y., Choi, K., Zhang, M., Zhou, Z., ... & Wu, B. (2008). Visual disability, visual function, and myopia among rural Chinese secondary school children: The Xichang Pediatric Refractive Error Study (X-PRES)—report 1. *Investigative Ophthalmology & Visual Science*, 49(7), 2888—2894.
- Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., & Zamora, P. (2013). Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *The Quarterly Journal of Economics*, 128(2), 531—580.
- Currie, J., & Almond, D. (2011). Human capital development before age five. In *Handbook of labor economics* (Vol. 4, pp. 1315–1486). Elsevier.
- Dallman, P. R. (1986). Biochemical basis for the manifestations of iron deficiency. *Annual Review of Nutrition*, 6(1), 13—40.
- De Benoist, B., Cogswell, M., Egli, I., & McLean, E. (2008). Worldwide prevalence of anaemia 1993—2005. WHO Global Database of anaemia.
- De Janvry, A., & E. Sadoulet. (2004). Conditional cash transfer programs: Are they really magic bullets?. *ARE Update*, 7(6), 9—11.
- Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2013). Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. *Education Policy in Developing Countries*, 285—338. Chicago: University of Chicago Press.
- Duflo, E. (2006). Field experiments in development economics. *Econometric Society Monographs*, 42, 322.
- Duflo, E., & Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, 118(3), 815—842.
- Duflo, E., Dupas, P., & Kremer, M. (2017). The impact of free secondary education: Experimental evidence from Ghana. Massachusetts Institute of Technology Working Paper. Cambridge, MA.
- Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4, 3895—3962.
- Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4), 1241—1278.
- Dupas, P. (2014). Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica*, 82(1), 197—228.
- Dupas, P., & Miguel, E. (2017). Impacts and determinants of health levels in low-income countries. *Handbook of economic field experiments*, (2), 3, 3—93.
- Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., ... & Emery, D. (2007). Effectiveness of reading and mathematics software products: Findings from the first student cohort.
- Engle, P. L., Black, M. M., Behrman, J. R., De Mello, M. C., Gertler, P. J., Kapiriri, L., ... & International Child Development Steering Group. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet*, 369(9557), 229—242.
- Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). Education technology: An evidence-based review (No. w23744). National Bureau of Economic Research.
- Evans, D. K., & Popova, A. (2016). Cost-effectiveness analysis in development: Accounting for local costs and noisy impacts. *World Development*, 77, 262—276.
- Faber, B., Sanchis-Guarner, R., & Weinhardt, F. (2015). ICT and education: Evidence from student home addresses (No. w21306). National Bur-

- eau of Economic Research.
- Fairlie, R. W., & Jonathan R. (2013). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. *American Economic Journal: Applied Economics*, 5(3), 211—240.
- Filmer, D., & N. Schady. (2009). Are there diminishing returns to transfer size in conditional cash transfers? World Bank Policy Research Working Paper No. 4999. Washington, DC: World Bank.
- Freire-Serén, M. J. (2001). Human capital accumulation and economic growth. *Investigaciones Economicas*, 25(3), 585—602.
- Fuchs, T., & Woosmann, L. (2004). Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at Schools. *Brussels Economic Review*, 47, 359—585.
- Gertler, P. J., Heckman, J. J., Pinto, R. R., Zanolini, A., Vermeerch, C., & Walker, S., ... & Grantham-McGregor, S. (2013). Labor market returns to early childhood stimulation: A 20-year followup to an experimental intervention in Jamaica. The World Bank.
- Girard, T., & Dion, P. (2010). Validating the search, experience, and credence product classification framework. *Journal of Business Research*, 63(9-10), 1079—1087.
- Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2, 945—1017.
- Glewwe, P., Park, A., & Zhao, M. (2016). A better vision for development: Eyeglasses and academic performance in rural primary schools in China. *Journal of Development Economics*, 122, 170—182.
- Goolsbee, A., & Guryan, J. (2006). The impact of internet subsidies in public schools. *The Review of Economics and Statistics*, 88(2), 336—347.
- Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., & International Child Development Steering Group. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, 369(9555), 60—70.
- Gray, R., & McCormick, M. C. (2005). Early childhood intervention programs in the US: Recent advances and future recommendations. *Journal of Primary Prevention*, 26(3), 259.
- Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2), 223—255.
- Guan, H., Wang, H., Du, K., Zhao, J., Boswell, M., Shi, Y., & Qian, Y. (2018a). The effect of providing free eyeglasses on children's mental health outcomes in China: A cluster-randomized controlled trial. *International Journal of Environmental Research and Public Health*, 15(12), 2749.
- Guan, H., Wang, H., Huang, J., Du, K., Zhao, J., Boswell, M., ... Rozelle, S. (2018). Health seeking behavior among rural left-behind children: Evidence from Shaanxi and Gansu provinces in China. *International Journal of Environmental Research and Public Health*, 15(5), 883.
- Guo, S., & Fraser, M. W. (2014). Propensity score analysis: Statistical methods and applications (Vol. 11). SAGE publications.
- Halterman, J. S., Kaczorowski, J. M., Aligne, C. A., Auinger, P., & Szilagyi, P. G. (2001). Iron deficiency and cognitive achievement among school-aged children and adolescents in the United States. *Pediatrics*, 107(6), 1381—1386.
- He, F., Linden, L., & MacLeod, M. (2008). How to teach english in India: Testing the relative productivity of instruction methods within the Pratham english language education program. New York: Columbia University. Mimeographed document.
- He, M., Huang, W., Zheng, Y., Huang, L., & Ellwein, L. B. (2007). Refractive error and visual impairment in school children in rural southern China. *Ophthalmology*, 114(2), 374—382.
- He, M., Zeng, J., Liu, Y., Xu, J., Pokharel, G. P., & Ellwein, L. B. (2004). Refractive error and visual impairment in urban children in southern China. *Investigative Ophthalmology & Visual Science*, 45(3), 793—799.
- Heckman, J. J., Lochner, L. J., & Todd, P. E. (2003). Fifty years of Mincer earnings regressions (No. w9732). National Bureau of Economic Research.
- Heckman, J. J., & Vytlacil, E. J. (2008). Econometrics evaluation of social program part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effect in new environment. *Handbook of econometrics*, 6, 4875—5143.
- Heckman, J. J. (1992). Randomization and social policy evaluation. In *Evaluating welfare and training programs*, ed. C Manski, I Garfinkel, pp. 201—30. Cambridge, MA: Harvard Univ. Press
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, 54(1), 3—56.
- Heckman, J. J., & Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition. Nber Working Papers.
- Heckman, J. J., Lochner, L., & Taber, C. (1998). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics*, 1(1), 1—58.
- Heinrich, C. J. (2007). Demand and supply-side determinants of conditional cash transfer program effectiveness. *World Development*, 35(1),

121—143.

- Horton, R. (2008). Maternal and child undernutrition: An urgent opportunity. *The Lancet*, 371(9608), 179.
- Hwang, G. J., Wu, P. H., & Chen, C. C. (2012). An online game approach for improving students' learning performance in web-based problem-solving activities. *Computers & Education*, 59(4), 1246—1256.
- Iannotti, L. L., Tielsch, J. M., Black, M. M., & Black, R. E. (2006). Iron supplementation in early childhood: Health benefits and risks. *The American Journal of Clinical Nutrition*, 84(6), 1261—1276.
- Imbens, G. W., & Joshua, D. A. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467—75.
- Jin, X., Sun, Y., Jiang, F., Ma, J., Morgan, C., & Shen, X. (2007). "Care for Development" intervention in rural China: A prospective follow-up study. *Journal of Developmental & Behavioral Pediatrics*, 28(3), 213—218.
- Kleiman-Weiner, M., Luo, R., Zhang, L., Shi, Y., Medina, A., & Rozelle, S. (2013). Eggs versus chewable vitamins: Which intervention can increase nutrition and test scores in rural China? *China Economic Review*, 24, 165—176.
- Kremer, M., & Glennerster, R. (2011). Improving health in developing countries: Evidence from randomized evaluations. *Handbook of health economics*, (2), 201—315.
- Lai, F., Luo, R., Zhang, L., Huang, X., & Rozelle, S. (2015a). Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. *Economics of Education Review*, 47, 34—48.
- Lai, F., Zhang, L., Bai, Y., Liu, C., Shi, Y., Chang, F., & Rozelle, S. (2016). More is not always better: Evidence from a randomised experiment of computer-assisted learning in rural minority schools in Qinghai. *Journal of Development Effectiveness*, 8(4), 449—472.
- Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M., & Rozelle, S. (2013). Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi. *Journal of Development Effectiveness*, 5(2), 208—231.
- Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M., & Rozelle, S. (2015b). Teaching the language of wider communication, minority students, and overall educational performance: Evidence from a randomized experiment in Qinghai Province, China. *Economic Development and Cultural Change*, 63(4), 753—776.
- Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. *Developmental Psychology*, 42(4), 627.
- Li, F., Song, Y., Yi, H., Wei, J., Zhang, L., Shi, Y., Chu, J., Johnson N., Loyalka, P., & Rozelle, S. (2017). The impact of conditional cash transfers on the matriculation of junior high school students into rural China's high schools. *Journal of Development Effectiveness*, 9(1), 41—60.
- Li, H., Loyalka, P., Rozelle, S., Wu, B., & Xie, J. (2015). Unequal access to college in China: How far have poor, rural students been left behind? *The China Quarterly*, 221, 185—207.
- Li, L., Huang, L., Shi, Y., Luo, R., Yang, M., & Rozelle, S. (2018). Anemia and student's educational performance in rural Central China: Prevalence, correlates and impacts. *China Economic Review*, 51, 283—293.
- Li, L., Lam, J., Lu, Y., Ye, Y., Lam, D. S., Gao, Y., ... & Congdon, N. (2010). Attitudes of students, parents, and teachers toward glasses use in rural China. *Archives of Ophthalmology*, 128(6), 759—765.
- Li, L., Song, Y., Liu, X., Lu, B., Choi, K., Lam, D. S., ... & Congdon, N. (2008). Spectacle acceptance among secondary school students in rural China: The Xichang Pediatric Refractive Error Study (X-PRES)—report 5. *Investigative Ophthalmology & Visual Science*, 49(7), 2895—2902.
- Linden, L. L. (2008). Complement or substitute? The effect of technology on student achievement in India. InfoDev.
- Liu, C., Zhang, L., Luo, R., Rozelle, S., Sharbono, B., & Shi, Y. (2009). Development challenges, tuition barriers, and high school education in China. *Asia Pacific Journal of Education*, 29(4), 503—520.
- Liu, C., Zhang, L., Luo, R., Wang, X., Rozelle, S., Sharbono, B., Adams, J., Shi, Y., Yue, A., Li, H., & Glauben, T. (2011). Early commitment on financial aid and college decision making of poor students: Evidence from a randomized evaluation in rural China. *Economics of Education Review*, 30(4), 627—640.
- Liu C., Zhang, Linxiu., Luo R., Rozelle, S., & Loyalka, P. (2010). The effect of primary school mergers on academic performance of students in rural China. *Journal of Educational Development*, 30(6), 570—585.
- Lochner, L. (2011). Non-production benefits of education: Crime, health, and good citizenship (No. w16722). National Bureau of Economic Research.
- Loyalka, P., C. Liu, Y. Song, Y. Hongmei, X. Huang, J. Wei, S. Rozelle, Y. Shi, J. Chu, & S. Rozelle. (2013). Can information and counseling help students from poor rural areas go to high school? Evidence from China. *Journal of Comparative Economics*, 41(4), 1012—1025.
- Loyalka, P., Sylvia, S., Liu, C., Chu, J., & Shi, Y. (2019). Pay by design: Teacher performance pay design and the distribution of student achievement. *Journal of Labor Economics*, 37(3), 621—662.



- Lozoff, B., & Georgieff, M. K. (2006). Iron deficiency and brain development. *Seminars in Pediatric Neurology*, 13(3), 158—165.
- Lozoff, B., Jimenez, E., Hagen, J., Mollen, E., & Wolf, A. W. (2000). Poorer behavioral and developmental outcome more than 10 years after treatment for iron deficiency in infancy. *Pediatrics*, 105(4), e51—e51.
- Luo, R., Emmers, D., Warrinner, N., Rozelle, S., & Sylvia, S. (2019b). Using community health workers to deliver a scalable integrated parenting program in rural China: A cluster-randomized controlled trial. *Social Science & Medicine*, 239, 112545.
- Luo, R., Jia, F., Yue, A., Zhang, L., Lyu, Q., Shi, Y.,... & Rozelle, S. (2017a). Passive parenting and its association with early child development. *Early Child Development and Care*, 189(10), 1709—1723.
- Luo, R., Miller, G., Rozelle, S., Sylvia, S., & Vera-Hernández, M. (2019a). Can bureaucrats really be paid like ceos? Substitution between incentives and resources among school administrators in China. *Journal of the European Economic Association*.1-37.
- Luo, R., Shi, Y., Zhang, L., Liu, C., Rozelle, S. & Sharbono, B. (2009). Malnutrition in China's rural boarding schools: The case of primary schools in Shaanxi province. *Asia Pacific Journal of Education*, 29(4), 481—501.
- Luo, R., Shi, Y., Zhang, L., Liu, C., Rozelle, S., Sharbono, B., ... & Martorell, R. (2012a). Nutrition and educational performance in rural China's elementary schools: Results of a randomized control trial in Shaanxi Province. *Economic Development and Cultural Change*, 60(4), 735—772.
- Luo, R., Shi, Y., Zhang, L., Zhang, H., Miller, G., Medina, A., & Rozelle, S. (2012b). The limits of health and nutrition education: Evidence from three randomized-controlled trials in rural China. *CESifo Economic Studies*, 58(2), 385—404.
- Luo, R., Shi, Y., Zhou, H., Yue, A., Zhang, L., Sylvia, S., ... & Rozelle, S. (2014). Anemia and feeding practices among infants in rural Shaanxi province in China. *Nutrients*, 6(12), 5975—5991.
- Luo, R., Wang, X., Zhang, L., Liu, C., Shi, Y., Miller, G., ... & Martorell, R. (2011a). High anemia prevalence in western China. *Southeast Asian Journal of Tropical Medicine and Public Health*, 42(5), 1204.
- Luo, R., Yue, A., Zhou, H., Shi, Y., Zhang, L., Martorell, R., ... & Sylvia, S. (2017b). The effect of a micronutrient powder home fortification program on anemia and cognitive outcomes among young children in rural China: A cluster randomized trial. *BMC Public Health*, 17(1), 738.
- Luo, R., Zhang, L., Liu, C., Zhao, Q., Shi, Y., Miller, G., ... & Martorell, R. (2011). Anaemia among students of rural China's elementary schools: Prevalence and correlates in Ningxia and Qinghai's poor counties. *Journal of Health, Population, and Nutrition*, 29(5), 471.
- Ma, X., Congdon, N., Yi, H., Zhou, Z., Pang, X., Meltzer, M. E., ... & Rozelle, S. (2015). Safety of spectacles for children's vision: A cluster-randomized controlled trial. *American Journal of Ophthalmology*, 160(5), 897—904.
- Ma, X., Zhou, Z., Yi, H., Pang, X., Shi, Y., Chen, Q., ... & Liu, Y. (2014). Effect of providing free glasses on children's educational outcomes in China: Cluster randomized controlled trial. *BMJ*, 349, g5740.
- Ma, Y., Congdon, N., Shi, Y., Hogg, R., Medina, A., Boswell, M., ... & Iyer, M. (2018a). Effect of a local vision care center on eyeglasses use and school performance in rural China: A cluster randomized clinical trial. *JAMA Ophthalmology*, 136(7), 731—737.
- Ma, Y., Gao, Y., Wang, Y., Li, H., Ma, L., Jing, J., ... & Congdon, N. (2018b). Impact of a local vision care center on glasses ownership and wearing behavior in northwestern rural China: A cluster-randomized controlled trial. *International Journal of Environmental Research and Public Health*, 15(12), 2783.
- Malcolm, L. A. (1970). Growth retardation in a New Guinea boarding school and its response to supplementary feeding. *British Journal of Nutrition*, 24(1), 297—305.
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2), 407—437.
- Miguel, Edward, and Michael Kremer. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1), 159—217.
- Miller, G., Luo, R., Zhang, L., Sylvia, S., Shi, Y., Foo, P.,... & Rozelle, S. (2012). Effectiveness of provider incentives for anaemia reduction in rural China: A cluster randomised trial. *BMJ*, 345, e4809.
- Mo, D., Luo, R., Liu, C., Zhang, H., Zhang, L., Medina, A., & Rozelle, S. (2014b). Text messaging and its impacts on the health and education of the poor: Evidence from a field experiment in rural China. *World Development*, 64, 766—780.
- Mo, D., Swinnen, J., Zhang, L., Yi, H., Qu, Q., Boswell, M., & Rozelle, S. (2013b). Can one-to-one computing narrow the digital divide and the educational gap in China? The case of Beijing migrant schools. *World Development*, 46, 14—29.
- Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., Qiao, Y., Boswell, M., & Rozelle, S. (2014a). Integrating computer-assisted learning into a regular curriculum: Evidence from a randomised experiment in rural schools in Shaanxi. *Journal of Development Effectiveness*, 6(3),

300—323.

- Mo, D., Zhang, L., Yi, H., Luo, R., Rozelle, S. & Brinton, C. (2013a). School dropouts and conditional cash transfers: Evidence from a randomized controlled trial in rural China's junior high schools. *Journal of Development Studies*, 49(20), 190—207.
- Moswela, B. (2006). Boarding schools as perpetrators of students' behavior problems. *Journal of Social Science*, 13(1), 37—41.
- Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1), 39—77.
- Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2016). Building state capacity: Evidence from biometric smartcards in India. *American Economic Review*, 106(10), 2895—2929.
- Mushkin, S. J. (1962). Health as an investment. *Journal of Political Economy*, 70(5, Part 2), 129—157.
- Nazmul, C., & D. Parajuli. (2008). Conditional cash transfers and female schooling: The impact of the female school stipend program on public school enrollments in Punjab, Pakistan. *Journal of Applied Economics*, 42(28), 3565—3583.
- Nie, J., Pang, X., Sylvia, S., Wang, L., & Rozelle, S. (2018). Seeing is believing: Experimental evidence on the impact of eyeglasses on academic performance, aspirations and dropout among junior high school students in rural China. *Economic Development and Cultural Change*. Retrieved from <https://www.journals.uchicago.edu/doi/10.1086/700631>.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects?. *Educational Evaluation and Policy Analysis*, 26(3), 237—257.
- Park, A., Shi, X., Hsieh, C. T., & An, X. (2015). Magnet high schools and academic performance in China: A regression discontinuity design. *Journal of Comparative Economics*, 43(4), 825—846.
- Park, A., Cai, F., & Du, Y. (2010). Can China meet its employment challenges?. Growing pains: Tensions and opportunity in China's transformation. The Walter H. Shorenstein Asia-Pacific Research Center, Stanford, 27—56.
- Priya Morjaria, Ian McCormick, & Clare Gilbert. (2019). Compliance and predictors of spectacle wear in schoolchildren and reasons for non-wear: A review of the literature. *Ophthalmic Epidemiology*, 1, 1—11.
- Resnikoff, S. (2008). Global magnitude of visual impairment caused by uncorrected refractive errors in 2004. *Bulletin of the World Health Organization*, 86(1), 63—70.
- Reynolds, A. J., Temple, J. A., & Ou, S. R. (2010). Preschool education, educational attainment, and crime prevention: Contributions of cognitive and noncognitive skills. *Children & Youth Services Review*, 32(8), 1054—1063.
- Richardson, J. T., & Long, G. L. (2003). Academic engagement and perceptions of quality in distance education. *Open Learning: The Journal of Open, Distance and e-Learning*, 18(3), 223—244.
- Ritchie, H., & Newby, T. J. (1989). Instruction: Classroom lecture/discussion vs live televised instruction: A comparison of effects on student performance, attitude, and interaction. *American Journal of Distance Education*, 3(3), 36—45.
- Rivera, J. A., Sotres-Alvarez, D., Habicht, J. P., Shamah, T., & Villalpando, S. (2004). Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children: A randomized effectiveness study. *Jama*, 291(21), 2563—2570.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247—252.
- Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a “scientifically based” reading program. *Economics of Education Review*, 23(4), 323—338.
- Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57—67.
- Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, 74(1), 199—250.
- Schultz, T. W. (1973). The value of children: An economic perspective. *Journal of Political Economy*, 81(2, Part 2), S2—S13.
- Sharma, A., Li, L., Song, Y., Choi, K., Lam, D. S., Zhang, M., ... & Congdon, N. (2008). Strategies to improve the accuracy of vision measurement by teachers in rural Chinese secondary school children: Xichang Pediatric Refractive Error Study (X-PRES) report no. 6. *Archives of Ophthalmology*, 126(10), 1434—1440.
- Sheng, X., Wang, J., Li, F., Ouyang, F., & Ma, J. (2019). Effects of dietary intervention on vitamin B 12 status and cognitive level of 18-month-old toddlers in high-poverty areas: A cluster-randomized controlled trial. *BMC Pediatrics*, 19(1), 1—9.
- Shi, L., Zhang, J., Wang, Y., Caulfield, L. E., & Guyer, B. (2010). Effectiveness of an educational intervention on complementary feeding practices and growth in rural China: A cluster randomised controlled trial. *Public Health Nutrition*, 13(4), 556—565.

- Shi, Y., Zhang, L., & Rozelle, S. (2015). When will we ever learn... to change policy: Current state of impact evaluation. *Journal of Development Effectiveness*, 7(4), 402—422.
- Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., ... & Eyers, J. (2015). Interventions for improving learning outcomes and access to education in low-and middle-income countries: A systematic review. 3ie Final Review. International Initiative For Impact Evaluation, London.
- Sun, K. T., Lin, Y. C., & Yu, C. J. (2008). A study on learning effect among different learning styles in a Web-based lab of science for elementary school students. *Computers & Education*, 50(4), 1411—1422.
- Sylvia, S., Luo, R., Zhang, L., Shi, Y., Medina, A., & Rozelle, S. (2013). Do you get what you pay for with school-based health programs? Evidence from a child nutrition experiment in rural China. *Economics of Education Review*, 37, 1—12.
- Sylvia, S., Ma, X., Shi, Y., Rozelle, S., & Lawell, C. Y. C. L. (2018). Ordeal mechanisms, information, and the cost-effectiveness of subsidies: Evidence from subsidized eyeglasses in rural China. C.Y. Cynthia Lin Lawell.
- Townsend, R. (2010). Financial structure and economic welfare: Applied general equilibrium development economics. *Annual Review of Economics*, 2(1), 507—546.
- US Department of Agriculture: Agricultural Research Service. (1999). USDA Nutrient Database for Standard Reference, Release 13. Nutrient Data Laboratory.
- Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pollitt, E., ... & International Child Development Steering Group. (2007). Child development: Risk factors for adverse outcomes in developing countries. *The Lancet*, 369(9556), 145—157.
- Walker, S. P., Wachs, T. D., Grantham-Mcgregor, S., Black, M. M., Nelson, C. A., & Huffman, S. L., et al. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, 378(9799), 1325—38.
- Wang, X., Congdon, N., Ma, Y., Hu, M., Zhou, Y., Liao, W., ... & Yi, H. (2017). Cluster-randomized controlled trial of the effects of free glasses on purchase of children's glasses in China: The PRICE (Potentiating Rural Investment in Children's Eyecare) study. *PloS One*, 12(11), e0187808.
- White, H. (2009). Theory-based impact evaluation: Principles and practice. *Journal of Development Effectiveness*, 1(3), 271—284.
- Wong, H. L., Shi, Y., Luo, R., Zhang, L., & Rozelle, S. (2014). Improving the health and education of elementary schoolchildren in rural China: Iron supplementation versus nutritional training for parents. *Journal of Development Studies*, 50(4), 502—519.
- World Health Organization. (2006). Sight test and glasses could dramatically improve the lives of 150 million people with poor vision.
- World Health Organization. (2014). Facts about blindness and visual impairment. Retrieved from <http://www.who.int/features/factfiles/blindness/en/>, 2014-8.pdf
- World Health Organization. (2011). Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. 2011. <http://apps.who.int/iris/handle/10665/85839>. Accessed 15 May 2016.
- Yi, H., Zhang, H., Ma, X., Zhang, L., Wang, X., Jin, L., ... & Rozelle, S. (2015). Impact of free glasses and a teacher incentive on children's use of eyeglasses: A cluster-randomized controlled trial. *American Journal of Ophthalmology*, 160(5), 889—896.
- Yue, A., Bai, Y., Shi, Y., Luo, R., Rozelle, S., Medina, A., & Sylvia, S. (2020). Parental migration and early childhood development in rural China. *Demography*, (57), 403—422.
- Yue, A., Shi, Y., Chang, F., Yang, C., Wang, H., Hongmei, Y. I., ... & Rozelle, S. (2014). Dormitory management and boarding students in China's rural primary schools. *China Agricultural Economic Review*.
- Yue, A., Shi, Y., Luo, R., Wang, B., Weber, A., Medina, A., ... & Rozelle, S. (2019). Stimulation and early child development in China: Caregiving at arm's length. *Journal of Developmental & Behavioral Pediatrics*, 40(6), 458—467.
- Zhang, L., Kleiman-Weiner, M., Luo, R., Shi, Y., Martorell, R., Medina, A., & Rozelle, S. (2013). Multiple micronutrient supplementation reduces anemia and anxiety in rural China's elementary school children. *The Journal of Nutrition*, 143(5), 640—647.
- Zhang, T., & Minxia, Z. (2006). Universalizing nine-year compulsory education for poverty reduction in rural China. *International Review of Education*, 52(3—4), 261—286.
- Zhou, C., S. Sylvia, L. Zhang, R. Luo, H. Yi, C. Liu, Y. Shi, P. Loyalka, J. Chu, A. Medina, & Rozelle, S. (2015). China's left-behind children: Impact of parental migration on health, nutrition, and educational outcomes. *Health Affairs*, 34(11), 1964—1971.
- Zhou, H., Sun, S., Luo, R., Sylvia, S., Yue, A., Shi, Y., ... & Rozelle, S. (2016). Impact of text message reminders on caregivers' adherence to a home fortification program against child anemia in rural western China: A cluster-randomized controlled trial. *American Journal of Public Health*, 106(7), 1256—1262.



## 注 释:

①在随机分配完成后,其中的7所学校关闭了,研究人员发现关闭的学校在干预组和对照组分布均等,并不影响随机分配的有效性,因此最终样本包括57所干预学校和56所对照学校。

②“有效工作日”指当天上课时间和放学时间间隔超过5小时,且早晚合照里的学生人数不少于8人。

③统计功效:在假设检验中是指当备择假设(H1)为真时正确地拒绝零假设(H0)的概率。在给定的显著性水准下,功效分析可以用于计算给定效应值时所需的最小样本数;相反地,功效分析也可以用来计算给定样本数时所能检验到的最小效应值。参见Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* 2nd. 1988. ISBN 0-8058-0283-5。

④Fisher的工作可以参见朱迪亚·珀尔、达纳·麦肯齐《为什么:关于因果关系的新科学》(中信出版集团,2019年)一书。

⑤参见Yeh, R. W., Valsdottir, L. R., Yeh, M. W., Shen, C., Kramer, D. B., Strom, J. B., ... & Nallamotheu, B. K. (2018). Parachute use to prevent death and major trauma when jumping from aircraft: Randomized controlled trial. *BMJ*, 363, k5094。

⑥The 2019 Altmetric Top 100, 摘自<https://www.altmetric.com/top100/2019/>。

⑦这一部分可以参照实施科学(Implementation Science)的研究成果:Dong(Roman)Xu, James Hughes, Chun Hao, Jing Gu, Jinghua Li, Jing Liao. (2019). “Research designs and methods in implementation science.” Sun Yat-sen Global Health Institute Implementation Science Training Course I, DOI: 10.13140/RG.2.2.25402.34240. <https://www.researchgate.net/project/Implementation-research-in-China>。

⑧项目组同时要求班主任、学生处长分别列出班上的贫困生以验证家庭资产计算方式的有效性。只有这三种方法同时确认的学生才被列为贫困生。

⑨内部有效性强调,研究目的是为了确定因变量的变化只受自变量影响,以排除其他的不受控因素。换句话说,内部有效性表示的是研究设计在多大程度上可以将因变量的变化归因于自变量的变化。参见:Elizabethann O’ Sullivan, Cary R. Rassel. *Research Methods for Public Administrators*. Addison Wesley Longman. 1999: 56. ISBN 0801318505. 3rd。

⑩外部有效性是指研究结果的普遍适应性。具体说来,就是指我们可以在多大程度上将研究结果推广到其他人或环境中。参见:Elizabethann O’ Sullivan, Cary R. Rassel. *Research Methods for Public Administrators*. Addison Wesley Longman. 1999: 56. ISBN 0801318505. 3rd。

⑪2019年12月20日摘自<https://www.3ieimpact.org/about-us>。

⑫2019年12月20日摘自<https://www.iadb.org/en/office-strategic-planning-and-development-effectiveness/development-effectiveness>。

(责任编辑 胡 岩)

## Practices and Experiences of Randomized Controlled Trials in China's Targeted Poverty Alleviations in Education

Shi Yaojiang<sup>1</sup> Zhang Linxiu<sup>2</sup> Chang Fang<sup>1</sup> Liu Han<sup>1</sup> Tang Bin<sup>1</sup> Gao Qiufeng<sup>1</sup> Guan Hongyu<sup>1</sup>  
Nie Jingchun<sup>1</sup> Yang Jie<sup>1</sup> Bai Yu<sup>1</sup> Li Ying<sup>1</sup> Tang Lei<sup>1</sup> Yue Ai<sup>1</sup> Ru Tong<sup>1</sup>

(1. The Center for Experimental Economics in Education, Shaanxi Normal University, Xi'an, Shaanxi 710119, China;

2. The United Nations Environment Programme - International Ecosystem Management Partnership, Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract:** Since the late 20th century, educational research has continued to intersect with such disciplines as economics, medicine, psychology, and computer science. As typical multidisciplinary research methods, randomized controlled trials (RCTs) have gradually become a key step in making developmental policies worldwide, especially in the fields of education and social policy. Since 2003, the Abdul Latif Jameel Poverty Action Lab (J-PAL) and Innovations for Poverty Action (IPA) have conducted nearly 300 RCTs in education in 44 countries around the world, whose empirical findings have been utilized to facilitate educational policies in developing countries such as South Asia, Africa, and Latin America. Owing to their enormous contribution to global poverty reduction, the founders of J-PAL also won the 2019 Nobel Prize in Economics. Although a large number of international studies have corroborated the positive roles of RCTs in promoting educational output and raising the incomes of poor groups, similar research efforts in the field of targeted poverty alleviation in education in China underestimated. This paper is a comprehensive review of literature that involves more than 30 publicly available RCTs on targeted poverty alleviation in education from 2007 to 2019. It is found that the RCTs in education conducted in poor rural China focus on four areas: teachers and education, information technology and education, nutrition & health and education, as well as early childhood development. The overarching topic discussed in these four areas is the intergenerational trap of human capital in poor rural China, followed by the exploration of more cost-effective and feasible solutions to issues existing regarding rural education. We further summarize the development of applying RCTs in promoting quality rural education in poor areas in terms of its developmental pathway, status quo, and practical experiences, so as to provide operable ways for the implementers of rural education, data-based empirical support for the decision-makers, as well as effective strategies verified by RCTs for national designers of targeted poverty alleviation policy in education. In addition, this paper also introduces the application of quasi-experimental methods to facilitate making targeted poverty alleviation policies in education in situations where RCTs may not be applicable.

**Keywords:** randomized controlled trials; rural education; poor rural areas; targeted poverty alleviation in education