

附录

附录一、平均受教育年限构建文献

（一）文献汇总和平均受教育年限数据库介绍¹

1. 文献汇总

表 3-1 平均受教育年限方法-文献-数据库对应表

方法	文献
永续盘存法	Barro & Lee (1993)
	Barro & Lee (2001)
	Cohen & Soto (2007)
出生队列趋势外推法	Barro & Lee (2013)
	Barro & Lee (2015)
	De La Fuente & Doménech (2015)
出生队列迭代后推法	Lutz et al. (2007)
	Bauer et al. (2012)
	Goujon et al. (2016)
	Springer et al. (2019)

2. 数据库介绍

永续盘存法属于早期使用的方法，所构建的数据已不再受现有研究所喜爱，也已经没有组织或个人对原先数据进行更新。目前公开的，持续更新的数据库包括分别基于趋势外推法和迭代后推法的 Barro-Lee 教育成就数据库（Barro-Lee Educational Attainment Dataset）和维特根斯坦中心人力资本数据（Wittgenstein Centre Human Capital Data）。

其中，Barro-Lee 教育成就数据库是目前研究最常用的平均受教育年限数据来源，为根据 Barro & Lee (2013) 构造。目前最新的数据是 2021 年更新的数据，提供了 146 个国家从 1950 年至 2015 年按年龄和性别划分的平均受教育年限面板数据。

此外，维特根斯坦中心人力资本数据也提供了 185 个国家 1950-2015 年的按

¹ 我们在上面介绍了三种不同的基本方法，然后这些方法在各文献中的应用要远比基本方法要复杂，需要考虑各种细节。由于篇幅所限，并且由于方法的具体应用依然属于技术性的部分，太多的、不可缺失的公式加大了理解的难度，因此，本文并不对这些方法的具体细节、每篇文章的不同进行介绍。读者若有兴趣，可阅读原文，或可与本文作者联系，以获得由本文作者总结的、关于用于以上基本方法的研究的具体细节和说明（附录一）。后文同。

年龄和性别划分的教育获得分布面板数据，同时提供了根据教育获得计算的平均受教育年限面板数据。相对而言，维特根斯坦中心人力资本数据提供的国家更多，内容更多，年龄段也更长，但目前其在文献中的使用还较少，被接受度远不及 Barro-Lee 教育成就数据库。我们猜测，维特根斯坦中心人力资本数据在数据构建过程中过多的平滑、估算和预测是其数据使用受限的主要原因。

表 2-2 平均受教育年限数据库

	Barro-Lee 教育成就数据库	维特根斯坦中心人力资本数据
使用方法	趋势外推法（前推+后推）	迭代后推法
方法文献	Barro & Lee (2013)	Springer et al. (2019)
国家	146	185
时间	1950–2015	1950–2015（2020 年数据即将公布）
最近更新	2021	2024
包含内容	平均受教育年限	教育获得分布、平均受教育年限
年龄段	15-24、25-34、…55-64 岁	15-24、25-34、…100+岁
	15-64 岁；25-64 岁	15+；25+等
性别	男女；男；女	男女；男；女
网网址	http://barrolee.com/ https://github.com/barrolee/BarroLeeDataSet	https://dataexplorer.wittgensteincentre.org/
被接受度	高	低

（二）永续盘存法：Barro & Lee (2001)

Barro 和 Lee 是较早使用永续盘存法构建各国平均受教育年限数据库的学者（Barro & Lee 1993；Barro & Lee 1996；Barro & Lee, 2001）。在具体实施中，他们具体采用如下公式来计算由于死亡和新增入学人口导致的各教育阶段人口变化：

$$H_{noedu,t} = H_{noedu,t-5}(1 - \delta_t) + L15_t * (1 - PRI_{t-5})$$

$$H_{1,pri} = H_{pri,t-5}(1 - \delta_t) + L15_t * (PRI_{t-5} - SEC_{t-5})$$

$$H_{sec,t} = H_{sec,t-5}(1 - \delta_t) + L15_t * SEC_{t-5} - L20_t * HIGH_t$$

$$H_{ter,t} = H_{ter,t-5}(1 - \delta_t) + L20_t * HIGH_t$$

上述每个公式都由两个部分构成，第一个部分为各教育阶段的存留人口，由公式 $H_{l,t-5}(1 - \delta_t)$ 给出，其中 δ_t 为死亡率，通过公式计算而得：

$$\delta_t = (L15_t + L_{t-5} - L_t)/L_{t-5}$$

其中 L_{t-5} 为过去的人口， $L15_t$ 为 t 时期的 15-29 岁人口，其相对于 $t - 5$ 期属

于新增的人口，两者相加代表不存在死亡时 t 时期的人口，故 $L15_t + L_{t-5} - L_t$ 为 t 时期相对于 $t-5$ 时变动（死亡）的人口。因此，死亡率 δ_t 为 $t-5$ 期到 t 期变动的人口（ $L15_t + L_{t-5} - L_t$ ）除以基期时的人口（ L_{t-5} ）。

第二个部分是新增受教育人口，由人口（ $L20_t$ 为20-24岁的人口）和入学率信息（ PRI_{t-5} ， SEC_{t-5} 和 $HIGH_t$ 为三个阶段教育的入学率）计算获得。不同的公式都是为了获得该教育阶段新增的人口²。

如方法部分所述，应用该方法存在的最大问题是其假设死亡率在各出生队列和教育阶段都是一样的，然后现实往往是教育程度越高其死亡率越低、年长的人死亡率越大。因此，有学者认为这是 Barro & Lee (2001) 的数据在时间上表现不稳定的重要原因（Cohen & Soto, 2007）。

文章进一步通过完成率将小学、中学、大学分成了完成和未完成两种状态。完成率的估计来自 Barro & Lee (1993)。

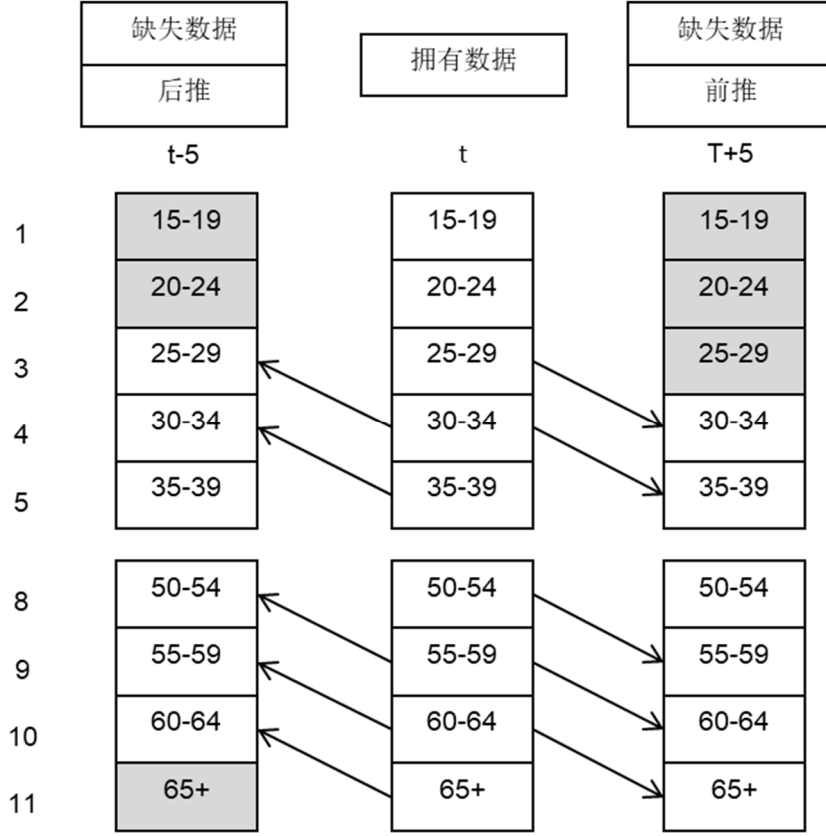
（三）出生队列趋势外推法：Cohen & Soto (2007)

Cohen & Soto (2007) 认为利用出生队列的信息可以解决死亡率在各出生队列都一致存在的问题，其率先将出生队列信息应用在平均受教育年限数据的构建工作中，提出了出生队列趋势外推法。

出生队列趋势外推法其实就是利用认为相同出生队列在不同年份时的信息是一样的，以此通过有数据年份的出生队列信息来推测无数据年份的出生队列信息。

为了简述 Cohen & Soto (2007)所采用的这一方法，我们给出了以下示意图：

² 需要再次提醒的是，获得的是最高学历达到该教育阶段的新增人口，那些获得了更高的学历的人不包含在内。比如，若 300 名新增人口完成了小学教育，其中 100 名又继续完成了初中教育，此式，小学教育的新增人口为 200（300-100）名。各公式的这一部分都是为了获得最高学历达到该阶段的新增人口。



具体而言，在利用趋势外推法时，其根据缺失数据年份的位置，选择年龄趋势前推和后推来得到 ys_t^a ：

$$ys_{t+5}^{a+1} = ys_t^a \quad \text{for } a = 3, \dots, 10$$

$$ys_{t-5}^{a-1} = ys_t^a \quad \text{for } a = 4, \dots, 11$$

其次，由于在前推中 ys_{t+5}^1 、 ys_{t+5}^2 和 ys_{t+5}^3 是无法获得的，其中 ys_{t+5}^1 是因为无对应年龄段数据， ys_{t+5}^2 和 ys_{t+5}^3 是因为 t 期的 15-19 和 20-24 年龄段人口（部分）还在接受教育，因此这一年龄段的平均受教育年限到 t+1 期肯定是会发展变化的。同样在后推中，也会有三个出生队列无法获得： ys_{t-5}^1 、 ys_{t-5}^2 、 ys_{t-5}^{11} 。

对于这些出生队列，Cohen & Soto (2007) 根据入学率数据推测获得：如果需要计算 1980 年时 60-64 岁教育获得，即 ys_{1980}^{10} 。假设进入初等教育的年龄为 6 岁，则该出生队列在 1922-1926 年间开始接受初等教育。通过计算 1924 年新进入小学一年级的人数与 6 岁人口的比率（即净入学率），可以得到该出生队列上小学的人口比例的估计。同样，按此方法可以估计其他教育阶段人口（即中等

教育和高等教育) 的占比³, 继而获得 ys_{1980}^{10} 。

最后, 由于其使用的趋势外推获得的是各出生队列的平均受教育年限, 因此 Cohen & Soto (2007) 基于公式 (4): $ys_t = \sum_{a=1}^{11} l_t^a ys_t^a$ 来计算各国的平均受教育年限。具体而言, 在获得了所有出生队列的 ys_t^a 后, 一国平均受教育年限公式则可以根据下式人口加总获得:

$$ys_{t-5} = l_{t-5}^1 ys_{t-5}^1 + l_{t-5}^2 ys_{t-5}^2 + \sum_{a=4}^{11} l_{t-5}^{a-1} ys_t^a + l_{t-5}^{11} ys_{t-5}^{11}$$

$$ys_{t+5} = l_{t+5}^1 ys_{t+5}^1 + l_{t+5}^2 ys_{t+5}^2 + l_{t+5}^3 ys_{t+5}^3 + \sum_{a=3}^{10} l_{t+5}^{a+1} ys_t^a$$

与使用永续盘存法的 Barro & Lee (2001) 相比, 由于使用出生队列外推法, Cohen & Soto (2007) 考虑到了各年龄人口的死亡率的差异。然而, Cohen & Soto (2007) 依然存在局限, 其虽然考虑到出生队列间死亡率的异质性, 但没有考虑到相同出生队列内不同教育阶段死亡率的异质性, 尤其是那些年龄较大的出生队列人口, 如 60-64 岁。正如之前所述, 受教育程度越高的人死亡率越低。某一出生队列中低受教育程度的更大死亡率和受教育程度更高的更小死亡率, 使得经过一段时间后, 该出生队列的平均受教育年限更大, 即 $ys_{t-5}^a < ys_t^{a+1}$; $ys_{t+5}^a > ys_t^{a-1}$ 。

对这一问题, Cohen & Soto (2007) 基于美国的数据估算对可能导致的误差进行了核算, 以说明这一情况不会对平均受教育年限数据的产生较大干扰: 在考虑了教育阶段和死亡率的关系之后, 美国最终平均受教育年限只提高了 2%, 这是一个较小的误差。

(四) 出生队列趋势外推法: Barro & Lee (2013)

Barro & Lee (2013) 同样使用了出生队列趋势外推法, 不同的是, Barro & Lee (2013) 在方法上有更多地细化。首先, 为了考虑年龄较大的出生队列内不同教育阶段死亡率的异质性, Barro & Lee (2013) 采用公式 (4) 和公式 (5) 在出生队列-教育阶段层面进行趋势外推:

$$ys_t = \sum_{a=1}^{13} l_t^a ys_t^a \quad \text{公式 (4)}$$

$$ys_t^a = \sum_{l=0}^l h_{l,t}^a Dur_{l,t} \quad \text{公式 (5)}$$

³ 由于每个类别的招生人数不能直接用于计算新进入人数(由于中继器、辍学和瞳孔的增长), 原文还队对其进行调整, 详情见原文。

也就是说 Barro & Lee (2013) 的出生队列趋势外推不再基于 ys_t^a ，而是基于 $h_{l,t}^a$ 。其次，Barro & Lee (2013) 首先考虑了更多的出生队列和更详细的教育阶段划分： $a = 11$ 表示 65-69 岁， $a = 12$ 表示 70-74 岁， $a = 13$ 表示 75 岁以上； $l = noedu$ ， $l = pri$ （完成，未完成）， $l = sec$ （完成，未完成）， $l = ter$ （完成，未完成）。

为了得到更加细化的教育阶段，插补程序划分需分三个步骤进行，第一步是考虑宽泛的教育阶段的人口占比；第二步是估计初等教育、中等教育和高等教育的完成率（*Completion Ratio*），通过完成率和各教育阶段的人口占比的乘积，将初等教育、中等教育和高等教育进一步细分为完成和未完成；第三步是计算平均受教育年限。

1. 广泛的各教育阶段的人口占比

与 Cohen & Soto (2007) 一致，Barro & Lee (2013) 认为在同一出生队列中，死亡率不随个体受教育阶段的变化而变化，即无论一个人的受教育程度如何，其生存率都是相同的。Barro & Lee (2013) 从现有人口普查信息中发现这一假设对 64 岁及以下人口是成立的，对老年群体并不成立，因此对 65 岁以上的死亡率进行调整。在这一情况下，Barro & Lee (2013) 的教育阶段人口占比的趋势外推公式如表 1 所示：

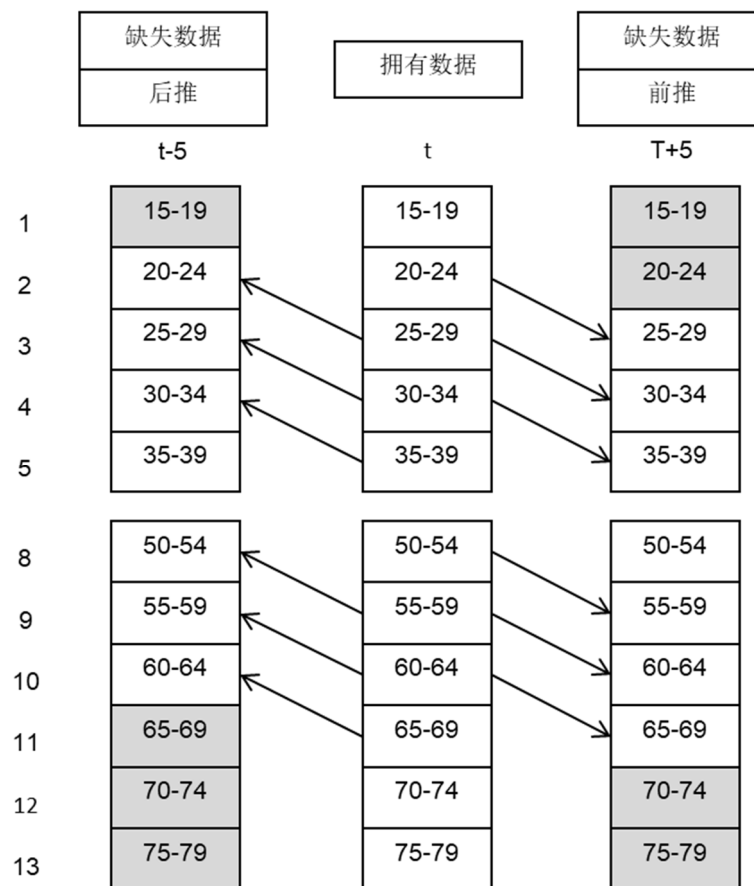
表 A1-1 各教育阶段人口占比的趋势外推公式

出生队列 (a)	后推	前推
15-19	$h_{l,t}^{15-19} = h_{l,t+5}^{15-19} - \Delta enroll_{l,t}^{15-19}$	$h_{j,t}^{15-19} = h_{l,t-5}^{15-19} - \Delta enroll_{l,t}^{15-19}$
20-24	$h_{l,t}^a = h_{l,t+5}^{a+1}$ ⁴	$h_{l,t}^{20-24} = h_{l,t-5}^{20-24} - \Delta enroll_{l,t}^{20-24}$
25-29, 30-34,..., 55-59, 60-64	$h_{l,t}^a = h_{l,t+5}^{a+1}$	$h_{l,t}^a = h_{l,t-5}^{a-1}$
65-69	$h_{l,t}^a = h_{l,t+5}^{a+1}(1 - \delta_l^a)^{-1}$	$h_{l,t}^a = h_{l,t-5}^{a-1}$
70-74, 75-79	$h_{l,t}^a = h_{l,t+5}^{a+1}(1 - \delta_l^a)^{-1}$	$h_{l,t}^a = h_{l,t-5}^{a-1}(1 - \delta_l^a)$

表中底纹为白色为直接外推的出生队列，底纹为灰的单元格为无法通过趋势外推补充、需进行额外处理操作的出生队列。也可以用上面同样的图形来表

⁴ 原文中，此处表格和文字不符，我们按照文字重新写了公式。

示。



首先，可见，在前推和后推的出生队列上，Barro & Lee (2013)和 Cohen & Soto (2007)有细微差别。在前推中，Cohen & Soto (2007)用 t 期的 25-29 推 $t+5$ 期的 30-34，而 Barro & Lee (2013) 则用 t 期的 20-24 推 $t+5$ 期的 25-29；在后推中 Cohen & Soto (2007)用 t 期的 30-34 推 $t-5$ 期的 25-29，而 Barro & Lee (2013) 则用 t 期的 25-29 推 $t-5$ 期的 20-24。同时，在高年龄段，Barro & Lee (2013)要比 Cohen & Soto (2007)多两个出生队列，并且在 Cohen & Soto (2007)中，无法外推获得的高出生队列主要是因为基期没有对应的更高的出生队列数据，而在 Barro & Lee (2013)中，是因为这些高年龄人口死亡率变动使得假设并不成立，无法进行外推。

此外，在这些无法通过趋势外推的出生队列数据的处理上，Barro & Lee (2013) 采取了与 Cohen & Soto (2007) 不同的处理方式：其进一步区分了低出生队列和高出生队列，分别采取不同的估算方法。在低出生队列上（15-19 岁、20-24 岁），Barro & Lee (2013) 同样采用入学率信息获得入学率调整因子（ $\Delta enroll_{i,t}^a$ ）。入学率调整因子（ $\Delta enroll_{i,t}^a$ ）为对应时期该教育阶段新增加的

人口占比由表 2 给出。

表 A1-2 各教育阶段入学率调整因子

教育阶段	以前推为例
未受教育	$-(enroll_{pri,t}^a - enroll_{pri,t-5}^a)$
初等教育	$(enroll_{pri,t}^a - enroll_{pri,t-5}^a) - (enroll_{sec,t}^a - enroll_{sec,t-5}^a)$
中等教育	$(enroll_{sec,t}^a - enroll_{sec,t-5}^a) - (enroll_{ter,t}^a - enroll_{ter,t-5}^a)$
高等教育	$(enroll_{ter,t}^a - enroll_{ter,t-5}^a)$

对 65 岁以上高出生队列利用死亡率 (δ_l^a) 进行调整。考虑到不同国家和不同教育类型死亡率的差别, 利用现有普查和调查数据, 通过使用加权最小二乘法估计下式:

$$h_{l,t}^a = (1 - \delta_l^a) * h_{l,t-5}^{a-1}, a = 70 - 74, 75 - 79; R = OECD, NOOECD; K = U, L$$

$$h_{l,t}^a = (1 - \delta_l^a)^2 * h_{l,t-10}^{a-2}, a = 75 - 79; R = OECD, NOOECD; K = U, L$$

分别获得 OECD 国家 ($R = OECD$) 和非 OECD 国家 ($R = NOOECD$), 低教育阶段 ($K = L$) 和高教育阶段 ($K = U$) 的死亡率调整因子。

此外, 在外推时, 虽然可以根据公式进行连续外推, 但限定最多外推两个时期, 即最多从 t 外推至 $t - 10$ 和 $t + 10$ 。

最后, 当同时可以从前推和后推获得某一年份的多个插补值时, 分别考虑 OECD 国家和非 OECD 国家, 通过现有的普查和调查数据, 使用加权最小二乘估计程序来估计下面的联立方程组获得前推和后推估计的权重:

$$h_{l,t}^a = \beta_{11}^R h_{l,t-5}^a + \beta_{21}^R h_{l,t+5}^a, \text{ where } \beta_{11}^R + \beta_{21}^R = 1$$

$$h_{l,t}^a = \beta_{12}^R h_{l,t-5}^a + \beta_{22}^R h_{l,t+10}^a, \text{ where } \beta_{12}^R + \beta_{22}^R = 1 \text{ and } \beta_{12}^R = \beta_{11}^R / (1 - \beta_{11}^R + \beta_{11}^{R^2})$$

$$\begin{aligned} h_{l,t}^a &= \beta_{13}^R h_{l,t-10}^a + \beta_{23}^R h_{l,t+5}^a, \text{ where } \beta_{13}^R + \beta_{23}^R = 1 \text{ and } \beta_{13}^R \\ &= \beta_{11}^{R^2} / (1 - \beta_{11}^R + \beta_{11}^{R^2}) \end{aligned}$$

$$h_{l,t}^a = \beta_{14}^R h_{l,t-10}^a + \beta_{24}^R h_{l,t+10}^a, \text{ where } \beta_{14}^R + \beta_{24}^R = 1 \text{ and } \beta_{14}^R = \beta_{11}^{R^2} / (1 - 2\beta_{11}^R + 2\beta_{11}^{R^2})$$

经过以上操作便可以获得广泛的各教育阶段人口的占比数据

2. 各教育阶段的完成率估计

初等教育、中等教育和高等教育的完成率同样基于出生队列趋势外推法，其中表 3 和表 4 分别给出了初等教育、中等教育和高等教育的完成率公式。

表 A1-3 初等教育、中等教育的完成率

出生队列 (a)	后推	前推
15-19	$w_{l,t}^{15-19} = w_{l,t+5}^{25-29} * (w_{l,t+5}^{15-19} / w_{l,t+5}^{25-29})$	$w_{l,t}^{15-19} = c_{l,t-5}^{20-24} * (w_{l,t-5}^{15-19} / w_{l,t-5}^{25-29})$
20-24	$w_{l,t}^a = w_{l,t+5}^{a+1}$	$w_{j,t}^{20-24} = c_{l,t-5}^{20-24} * (c_{l,t-5}^{20-24} / w_{l,t-5}^{25-29})$
25-29, ...70-74	$w_{l,t}^a = w_{l,t+5}^{a+1}$	$w_{l,t}^a = w_{l,t-5}^{a-1}$
75-79	$w_{l,t}^{75-79} = c_{l,t+5}^{75-79} * (c_{l,t+5}^{75-79} / c_{l,t+5}^{70-74})$	$w_{l,t}^{75-79} = sh_{l,t-5}^{70-74} * w_{l,t+5}^{70-74} + sh_{l,t-5}^{75-79} * w_{l,t+5}^{75-79}$

注： $sh_t^a = pop_t^a / pop_t^{70-79}$ ，也就是 t 时出生队列（ a ）的人口占 70-79 岁人口的比例，下同。

表 A1-4 高等教育的完成率

出生队列 (a)	后推	前推
15-19, 20-24	$c_{l,t}^a = c_{l,t+5}^a$	$c_{l,t}^a = c_{l,t-5}^a$
25-29	$c_{l,t}^a = c_{l,t+5}^{a+1}$	$c_{l,t}^{25-29} = c_{l,t-5}^{25-29} * (c_{l,t-5}^{25-29} / c_{l,t-5}^{30-34})$
30-34, ...70-74	$c_{l,t}^a = c_{l,t+5}^{a+1}$	$c_{l,t}^a = c_{l,t-5}^{a-1}$
75-79	$c_{l,t}^{75-79} = c_{l,t+5}^{75-79} * (c_{l,t+5}^{75-79} / c_{l,t+5}^{70-74})$	$c_{l,t}^{75-79} = sh_{l,t-5}^{70-74} * c_{l,t+5}^{70-74} + sh_{l,t-5}^{75-79} * c_{l,t+5}^{75-79}$

同样，当同时可以从前推和后推获得某一年份的多个插补值时，分别考虑 OECD 国家和非 OECD 国家，通过现有的普查和调查数据，使用加权最小二乘估计程序来估计下面的联立方程组获得前推和后推估计的权重：

$$c_{l,t}^a = \beta_{15}^R c_{l,t-5}^a + \beta_{25}^R c_{l,t+5}^a, \text{ where } \beta_{15}^R + \beta_{25}^R = 1$$

$$c_{l,t}^a = \beta_{16}^R c_{l,t-5}^a + \beta_{26}^R c_{j,t+10}^a, \text{ where } \beta_{16}^R + \beta_{26}^R = 1 \text{ and } \beta_{16}^R = \beta_{15}^R / (1 - \beta_{15}^R + \beta_{15}^{R^2})$$

$$c_{l,t}^a = \beta_{17}^R c_{l,t-10}^a + \beta_{27}^R c_{l,t+5}^a, \text{ where } \beta_{17}^R + \beta_{27}^R = 1 \text{ and } \beta_{17}^R = \beta_{11}^{R^2} / (1 - \beta_{11}^R + \beta_{11}^{R^2})$$

$$\begin{aligned} c_{l,t}^a &= \beta_{14}^R c_{l,t-10}^a + \beta_{24}^R c_{l,t+10}^a, \text{ where } \beta_{18}^R + \beta_{28}^R = 1 \text{ and } \beta_{18}^R \\ &= \beta_{15}^{R^2} / (1 - 2\beta_{15}^R + 2\beta_{15}^{R^2}) \end{aligned}$$

3. 最终的平均受教育年限计算

在使用 $c_{l,t}^a$ 和 $h_{l,t}^a$ 获得更细分类的 $h_{l,t}^a$ 后，使用公式（2）和公式（3）即可获得各国的平均受教育年限⁵。

在经过复杂的处理后，[Barro & Lee \(2013\)](#) 认为其提供的受教育程度的估计为广泛的国家提供了一个合理的人力资本存量的代理。事实上，[Barro & Lee \(2013\)](#) 所构造的数据也是先进国别研究中最常用的数据。

（五）出生队列迭代后推法：Springer et al. (2019)

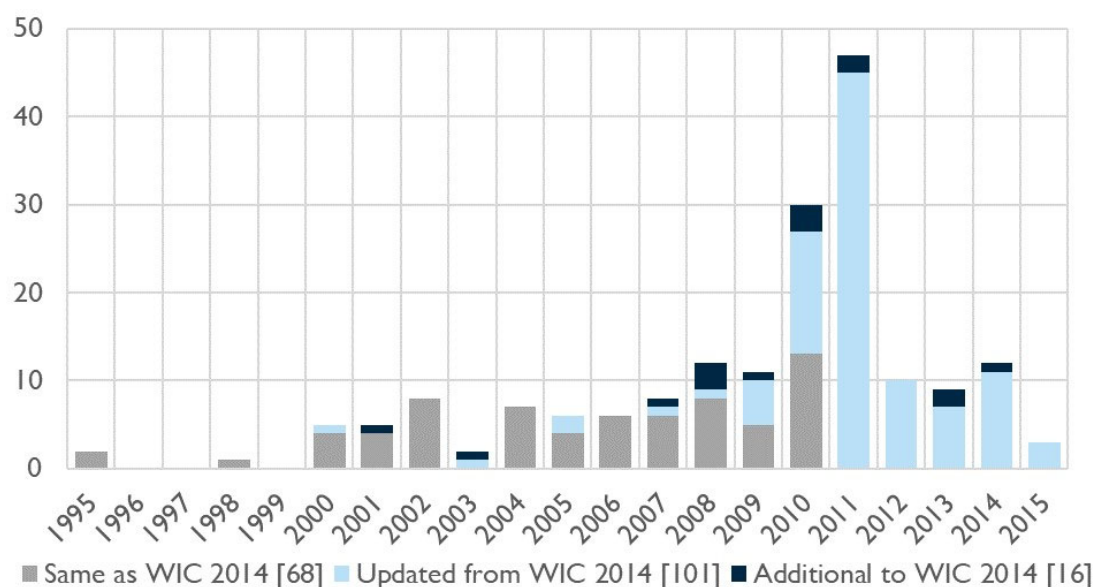
[Springer et al. \(2019\)](#) 是该系列文献最新的一篇。在方法部分介绍的步骤框架下，其进行了多项改良。下面我们将重新整理原文，并按方法部分介绍的步骤详细介绍这篇文章。

1. 数据收集和处理

（1）教育信息数据（基年+历史的）

基年的数据包括按出生队列-性别-教育阶段划分的人口（及占比）数据。基年的数据在地理覆盖范围方面，文章搜集整理的数据涵盖了了 185 个国家(占所有国家的 92%)，这一数据覆盖了全球 99% 的人口，使得该数据集成为国际上最全面的数据集。然后在这些国家中，只有 112 个国家的数据是 2010-2015 年的，由于数据的缺乏，一些国家依然只能依赖于早期的数据，如 [Central African Republic \(1995\)](#)、[Pakistan \(1998\)](#) 和 [Turkmenistan \(1995\)](#)。

⁵ Barro & Lee (2013) 基于同样的方法构造了男性和女性的平均受教育年限。见 Barro & Lee (2013) 以获得更多地细节。



Source: Springer et al. (2019)。

图 A1-1 基年数据年份示意图

对教育阶段进行标准化处理，按六个等级进行归类。由于只有 153 个国家拥有完整的六个等级的数据，其余国家需要估算，估算主要基于两种方式：（1）利用其他可替代的数据（比如拥有更详细的教育类别的劳动力调查（Labor Force survey, LFS）数据）的信息来替代，（2）用相似国家的信息进行类推。

表 A1-5 各教育阶段数据缺失情况概览

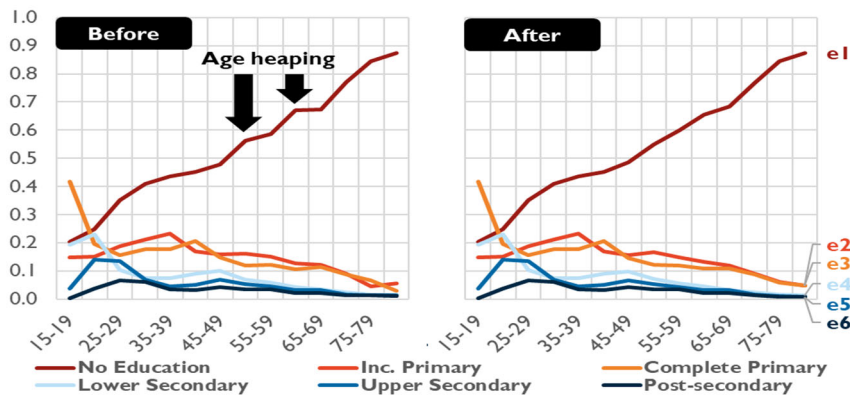
Case	No education	Incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	No. of countries
1	✓	✓	✓	✓	✓	✓	153
2	✓	►	✓	✓	✓	✓	14
3	►	►	✓	✓	✓	✓	8
4	✓	►	✓	►	✓	✓	5
5	►	✓	✓	✓	✓	✓	3
6	✓	►	►	✓	✓	✓	1
7	►	►	►	✓	✓	✓	1
N	173	156	183	180	185	185	185

Source: Springer et al. (2019)。

然后对发展中国家（尤其是撒哈拉以南非洲国家）出现的年龄跳跃（Age Heaping）进行处理。这些国家的人在报告年龄时因为四舍五入而误报，导致人口年龄分布在年龄尾数为 0 或 5 处出现聚集。这种误报对于教育程度低的人较为可能，对于受过较多教育的人可能性较低，因此，这种情况通常出现在 50 岁

和 60 岁。文章应用平滑技术来校正相关教育类别中的年龄堆积（如果四舍五入仅在一个年龄组中可见则利用函数或移动平均来处理），并对应地调整其他教育类别中的份额。同样，文章也使用相同的平滑技术来调整数据中的其他不合理的地方，例如不同出生队列的教育份额的不平滑趋势。

图 A1-2 年龄跳跃（Age Heaping）平滑处理示意图



Source: Springer et al. (2019)。

对于多来源的数据，其通过有效性检验（通过比较相同出生队列的人在不同时期的教育获得）来选择最好的基年数据。这种检验也是基于后推的基本假设：在理想的情况下，在低迁移人口、低死亡率、完成教育后的年龄组，教育阶段人口占比在不同时间也应当非常相似。

这其中，尤其是第二个问题，与之前研究不一样的是，该文想在数据延展过程中利用历史的教育信息数据。

历史教育信息数据主要是作为基年数据插补时的补充数据。因此这一信息时可有，并不是必须得。在文章的 185 个国家中，有 141 个国家拥有多个时间点的历史教育信息数据。⁶

需要注意的是，历史教育信息数据也需要和基年数据一样，进行一些标准化处理。

（A）对历史数据和基年数据：不同年龄教育分布的平滑和延展（the smoothing and extension of the age-specific education distribution）。

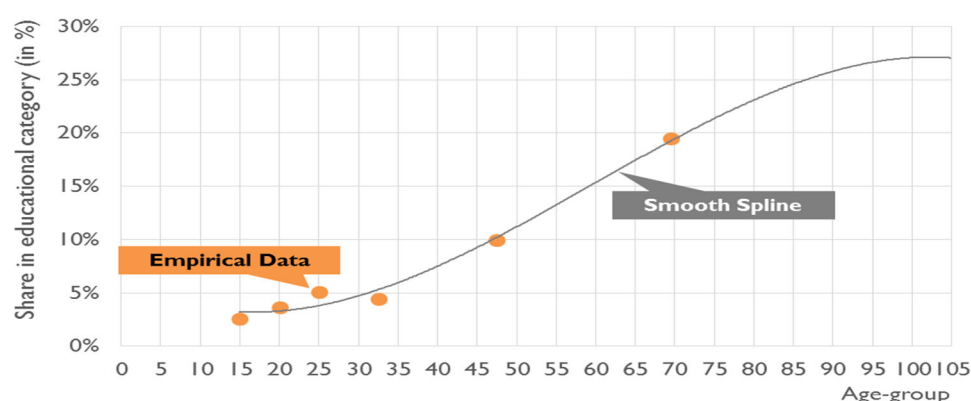
这一处理是为了对缺失教育成绩信息的年龄组进行插补，并将出生队列延伸至 105+。比如我们拥有每一国家每一年份的每一教育阶段的大部分出生队列

⁶ 对于这些数据可能存在的问题和处理办法见文章补充文件（supplementary file）。

人口（比如，15-19、。。。70+）占比数据。我们进行这一处理需要分成两个部分，第一个部分是使用的三次样条函数(Cubic Spline Function)，对 15-19：70+的数据进行平滑拟合插值；

第二个部分是使用 70+以下的 5 个年龄组，使用对数扩展（Logarithmic Extension）插值，使得最高年龄组为 105+。然而，需要指出的是，虽然最终数据只提供 100+的结果，但一个额外的年龄组是迭代方法必不可少的：在每一次迭代过程中，由于 t 期的教育分布被用来后推 t-5 期的教育分布，因此若 t 期的最高年龄组为 100+，那么 t-5 期就是 95-99，100+变空缺了。因此，需要多一个年龄组：105+。因此，经处理后，每一个时间点（历史数据和基年数据）的数据的年龄都是到 105+。下图给出了一个示例。

图 A1-3 某一年份各出生队列教育占比平滑示意图



Source: Springer et al. (2019)。

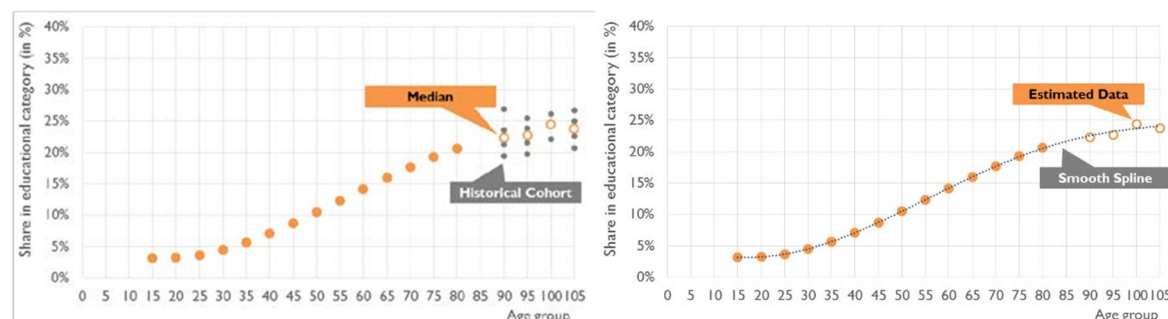
（B）对历史数据和基年数据：数据点移动至 0/5。这一处理是因为并非所有数据的年份都是以 0/5 结尾，因此需要将数据对齐至 0/5，比如将 1971 年的数据移动至 1970。在早先，包括 Berro & Lee (2013)、Goujon et al. (2016) 都只是简单的将 1971 年的数据当作 1970 年的数据。该文使用线性插补来实现这一数据时间点的移动，比如，如果对 2012 年进行时间点的调整，在获得其年龄组的平滑趋势后，如上图，其 20-24 年数据原为 20 对应的趋势线取值，调整后，为 18 对应的趋势线取值。

（C）只基年数据：当拥有相同出生队列历史教育信息时，延展数据时使用这些信息以提高延展数据的精确性。⁷这一步主要是对基年的数据。

⁷ 使用这些历史信息的前提是其可靠、可比且一致，因此需要对这些历史教育信息进行验证，并将教育类别不一致的数据点清空（比如，1985 年 75-79 岁的 Incomplete Primary）；然后利用反加总（Reverse Cumulative Sum：从中学后教育往前加总，和为 100）和反递进比（Reverse Education Progression Ratios：

如下图左侧所示，如果拥有这个出生队列多个历史教育信息，使用这些点的中位数作为这个出生队列的值。然后再如右图，进行平滑取值。

图 A1-4 基于历史数据的教育占比扩展示意图



Source: [Springer et al. \(2019\)](#)。

经过处理后，我们便有了用于迭代的基年的出生队列-性别-教育阶段人口占比数据。唯一的问题是，大部分国家的基年数据使用的都是 2015 年之前的，由于数据想获得 1950-2015 年的数据，这也意味着，很多国家的数据需要预测至 2015 年。

(2) 人口结构数据

按年龄（5 岁为一出生队列）和性别的 185 个国家 1950-2015 年（5 年间隔）的人口数据（必须）。

(3) 生命表（life table）数据和生存率

按性别划分的 185 个国家 1950-2015 年（5 年间隔）的生命表（Life table）数据。文中以上两项数据均来自 WPP2017⁸。

如方法部分所述，这一数据主要用于结合文章设定的各教育阶段预期寿命差信息，构建不同年份-出生队列-性别-教育阶段的生存率 $Survival Ratios(a - 5, l, t - 5, sex)$ 。WPP 提供的生命表数据中提供了各国的不同年份-年龄（出生队列）-性别的 15 岁预期寿命信息（e15），因此只需要添加教育阶段维度信息，便可以获得各国的不同年份-年龄（出生队列）-性别-教育阶段的 e15，继而获得

从高中到初中的比率）来分别对 No Education 和非 No Education 的数据点进行补值。这些具体操作详见 [Springer et al. \(2019\)](#)。

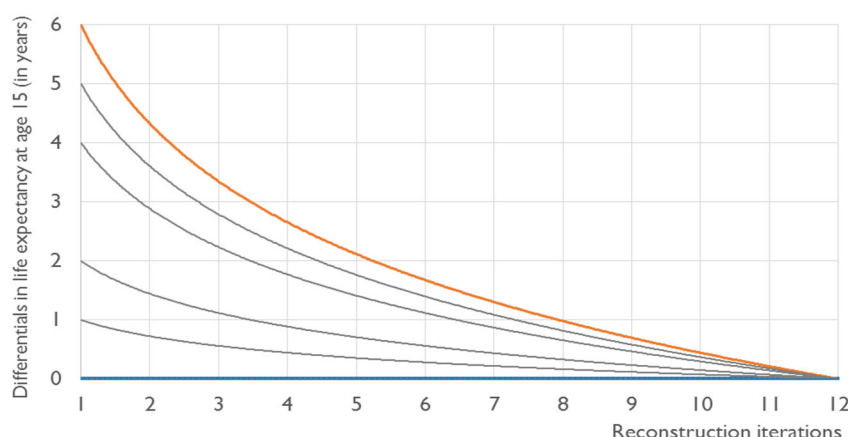
⁸ 由于 UN2017 年的数据在 1985-2015 年间提供了 100+ 的人口、预期寿命和死亡数据，但在 1950-1985 只提供了 85+ 的数据，因此文章需要对 1950-1985 年的数据进行插补，使得年份都到 100+。而最新的 WPP2024 (<https://population.un.org/wpp/>) 提供了所有年份 100+ 的数据，因此我们不对其插补的方法进行介绍。若想了解详见 [Springer et al. \(2019\)](#)。

Survival Ratios($a - 5, l, t - 5, sex$)。

教育阶段维度信息主要是各教育阶段预期寿命差信息。这一信息主要来自实证中的总结规律，并加一些设定。该文设定男性和女性的 15 岁时的预期寿命差在无教育和高中后教育之间分别为 6 年和 4 年⁹，并且假设 no education, some primary, completed primary, lower, upper, and post-secondary education levels 的男性和女性的 15 岁时的预期寿命差的比例分别为“1-1-2-1-1”（K.C. et al., 2015）。在之前的研究（Goujon et al., 2016）中，文章设定预期寿命差不随时间变化，该文对这一设定进行了改良，该文 15 岁时的预期寿命差的比例会以对数形式逐渐趋于无差异。这样设定的原因在于向零差异收敛更加符合流行病学的理论：在 20 世纪 50 年代，与目前占主导地位的慢性病相比，传染病更加流行，因此受教育程度所起的作用更小。

如下图所示，在 1（对应的 2010 年到 2005 年的迭代）处¹⁰，其预期寿命差比例为 1-1-2-1-1，在 12（对应的是 1955 年到 1950 年的迭代）处，已经无差异。

图 A1-5 15 岁时预期寿命(e15)迭代对数型收敛示意图



Source: Springer et al. (2019)。

2. 迭代计算

该文的并没有给出具体的迭代步骤，但列出了迭代过程的六个基本要素：

⁹ 在之前的数据构建中已经是这个设定了（即 2010 年的数据男女预期寿命差异已经是 6/4 年了），按理说 2015 年会有变化，但文中并没有说明，似乎还是沿用之前的。

¹⁰ 虽然是构建 1950-2015 年的数据，但文章多处还是以 2010 年数据的为例。

（1）后推

分性别，循着出生队列，将t期各出生队列的教育分布（人口占比）后推至t-5期：

（2）最高出生队列的空缺补值。

在本文中，由于t期的105+后推至t-5期的100-104，因而使得t-5期最高的出生队列105+空缺，需要补值。补值方法与教育信息数据中延展¹¹数据的方法一致。

（3）生存率调整

将上面获得的生存率信息应用于后推获得的t-5期数据上。

（4）学历提升调整。

学历提升（Education Transition, TR）调整限定在15-34岁，即34岁以上不会存在某一个人再进入学校学习以获得更高的学历。在各个教育阶段提升上，限定的年龄也不同。如下表，从incompleted primary提升至primary，即 TR_{e23} 只容许出现在低于15-19岁的群体上，而 TR_{e34} 、 TR_{e45} 、 TR_{e56} 分别限定为低于20-24岁、25-29岁、30-34岁的群体。也就是说，只有图中？的方格才有学历提升的可能。

表 A1-6 出生队列教育提升示意图

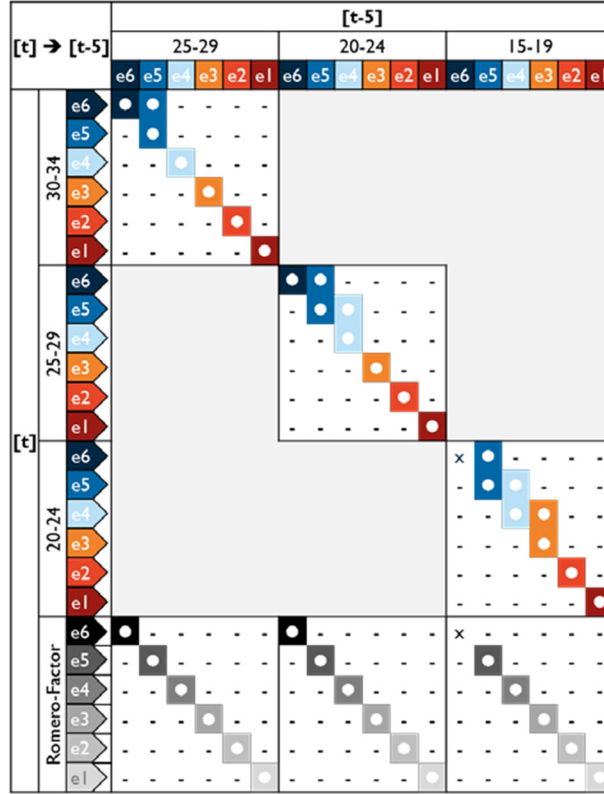
age	TR_{e12}	TR_{e23}	TR_{e34}	TR_{e45}	TR_{e56}
15--19	⊗	⊗	?	?	?
20--24	○	○	⊗	?	?
25--29	○	○	○	⊗	?
30--34	○	○	○	○	⊗

Source: Springer et al. (2019)。

下图给出了从T到T-5期可能存在的学历变化：分别为保持不变和下降至下一个教育阶段。

图 A1-6 相邻年份间不同出生队列学历提升示意图

¹¹ 对数扩展（Logarithmic Extension）插值和利用历史信息方法



Source: Springer et al. (2019)。

理想状态下，当我们拥有两个相连时间点的教育信息数据时，我们可以计算转换比例（Transition Ratios, TR）以从 e5 到 e6 为例：

$$TR_{e56}^{25 \rightarrow 30, t \rightarrow t+5} = \frac{P_{e6}^{30, t+5} - P_{e6}^{25, t}}{P_{e5}^{25, t}} = \frac{20\% - 5\%}{25\%} = 60\%$$

这一值衡量的是在 t 到 t+5 的过程中，有 60% 的人口将会从低教育阶段（e5）提升至高教育阶段（e6）。

然而需要注意的是，我们的迭代过程是一个后推过程，即从 t 到 t-5，因此我们需要使用的是保持不变的比率，称之为后推转换比率（Backward Transition Ratios, bTR）：

$$bTR_{e66}^{30 \rightarrow 25, t \rightarrow t-5} = \frac{P_{e6}^{25, t-5}}{P_{e6}^{30, t}} = \frac{5\%}{20\%} = 25\%$$

这一值衡量的是在后退过程中，有多少比例的人口依然保持有原有的教育阶段。而那些改变了教育阶段人口就是在 t-5 到 t 过程中，从低教育阶段（e5）提升至高教育阶段（e6）的人口。

以上说明，我们需要计算的是**bTR**以供后推时进行学历提升调整。

然而，事实上，我们是基于基年的数据进行后推。也就是说，我们只有一个时间点的教育数据是已知的。这个时候我们需要从这时间点的教育数据估算**bTR**。

估算过程大致包含以下几步：

(1) 反方向加总，获得大于等于这一教育等级的占比：

$$P_{l+}^a = \sum_l^6 P_l^a$$

比如，以文中例子：

表 A1-7 实例数据的 15-34 岁各阶段教育分布

Time	Age	P_{e1}	P_{e2}	P_{e3}	P_{e4}	P_{e5}	P_{e6}
t	15-19	0.1%	0.3%	2.0%	77.3%	20.4%	0.0%
t	20-24	0.1%	0.5%	1.3%	25.6%	64.1%	8.5%
t	25-29	0.1%	0.7%	2.0%	27.5%	53.1%	16.7%
t	30-34	0.4%	0.7%	2.7%	31.5%	45.5%	19.2%

Source: Springer et al. (2019)。

这一步将会产生下表结果：

表 A1-8 step1 反向累积

Time	Age	P_{e1+}	P_{e2+}	P_{e3+}	P_{e4+}	P_{e5+}	P_{e6}
t	15-19	100.0%	99.9%	99.6%	97.6%	20.4%	0.0%
t	20-24	100.0%	99.9%	99.4%	98.1%	72.6%	8.5%
t	25-29	100.0%	99.9%	99.1%	97.2%	69.7%	16.7%
t	30-34	100.0%	99.6%	98.9%	96.2%	64.7%	19.2%

Source: Springer et al. (2019)。

然后 Education Attainment Progression Ratio (EAPR)分别通过 $e2+/e1+$ 、 $e3+/e2+$ 、 $e4+/e3+$ 、 $e5+/e4+$ 、 $e6+/e5+$ 计算而得：

$$EAPR_{l \rightarrow l+1}^a = \frac{P_{l+1+}^a}{P_{l+}^a}$$

这一步得到的是，比如，在高中以上学历的总人口中，大学以上学历的人口占比，对应了入学率这一概念。以文中的例子，这一步会得到以下结果：

表 A1-9 step2 教育成就递进率 (Education Attainment Progression Ratio)

Time	Age	EAPR12	EAPR23	EAPR34	EAPR45	EAPR56
t	15-19	0.999	0.997	0.980	0.209	0.000
t	20-24	0.999	0.995	0.987	0.740	0.116
t	25-29	0.999	0.993	0.980	0.717	0.239
t	30-34	0.996	0.993	0.973	0.673	0.297

Source: [Springer et al. \(2019\)](#)。

然后取logit形式：

$$\text{logitEAPR}_{l \rightarrow l+1} = \log \left(\frac{\text{EAPR}_{l \rightarrow l+1}}{1 - \text{EAPR}_{l \rightarrow l+1}} \right)$$

这一结果衡量的是高中以上学历的口中，上了大学，获得大学以上学历的人口与没上大学，依然是高中学历的人口之比的对数。

得到以下结果：

表 A1-10 step3 教育成就递进率的逻辑形式

Time	Age	<i>logitEAPR12</i>	<i>logitEAPR23</i>	<i>logitEAPR34</i>	<i>logitEAPR45</i>	<i>logitEAPR56</i>
t	15-19	7.344	5.676	3.911	-1.333	-9.924
t	20-24	7.014	5.344	4.328	1.043	-2.026
t	25-29	6.587	4.915	3.908	0.931	-1.159
t	30-34	5.406	5.005	3.569	0.72	-0.861

Source: [Springer et al. \(2019\)](#)。

然后利用线性方程对 $\text{logitEAPR}_{l \rightarrow l+1}$ 进行外推插补¹²，获得：

表 A1-11 step4 外推插补

Time	Age	<i>logitEAPR12</i>	<i>logitEAPR23</i>	<i>logitEAPR34</i>	<i>logitEAPR45</i>	<i>logitEAPR56</i>
t	15-19	6.907	5.516	4.250	0.863	-0.861
t+5	20-24		5.517	4.495	1.025	-0.824
t+10	25-29		5.806	4.817	1.183	-0.786
t+15	30-34		5.806	5.107	1.344	-0.751

注：本表为文中数据倒推得到。

再将插补得到的 $\text{ogitEAPR}_{l \rightarrow l+1}$ 转换为 $\text{EAPR}_{l \rightarrow l+1}$ ：

表 A1-12 step4 外推插补结果的教育成就递进率

Time	Age	<i>EAPR12</i>	<i>EAPR23</i>	<i>EAPR34</i>	<i>EAPR45</i>	<i>EAPR56</i>
t	15-19	0.999	0.996	0.986	0.703	0.297

¹² 对什么插补，如何插补都是未知。

t+5	20-24	1.000	0.996	0.989	0.736	0.305
t+10	25-29	1.000	0.997	0.992	0.765	0.313
t+15	30-34	1.000	0.997	0.994	0.793	0.321

注：本表为文中数据倒推得到。

以便进一步获得 t+5、t+10 的出生队列 a 的插补加总分布（extrapolated cumulative educational distribution） eP_{t+}^a ，获得：

表 A1-13 step4 外推插补结果的反向累积

Time	Age	eP_{e1+}	eP_{e2+}	eP_{e3+}	eP_{e4+}	eP_{e5+}	eP_{e6}
t	15-19	100.00%	99.90%	99.50%	98.10%	69.00%	20.50%
t+5	20-24	100.00%	100.00%	99.60%	98.50%	72.50%	22.10%
t+10	25-29	100.00%	100.00%	99.70%	98.90%	75.70%	23.70%
t+15	30-34	100.00%	100.00%	99.70%	99.10%	78.60%	25.20%

Source: Springer et al. (2019)。

然后应用公式计算：

表 A1-14 后推转换比率

age	bTR_{e22}	bTR_{e33}	bTR_{e44}	bTR_{e55}	bTR_{e66}
15--19	⊗	⊗	$bTR_{e44}^{20 \rightarrow 15}$	$bTR_{e55}^{20 \rightarrow 15}$	○
20--24	○	○	⊗	$bTR_{e55}^{25 \rightarrow 20}$	$bTR_{e66}^{25 \rightarrow 20}$
25--29	○	○	○	⊗	$bTR_{e66}^{30 \rightarrow 25}$
30--34	○	○	○	○	⊗

Source: Springer et al. (2019)。

其中：

$$bTR_{e66}^{30 \rightarrow 25, t+5 \rightarrow t} = \frac{P_{e6}^{25, t}}{eP_{e6}^{30, t+5}}$$

$$bTR_{e66}^{25 \rightarrow 20, t+5 \rightarrow t} = \frac{P_{e6}^{20, t}}{(eP_{e6}^{30, t+10} * bTR_{e66}^{30 \rightarrow 25, t+5 \rightarrow t})}$$

$$bTR_{e55}^{25 \rightarrow 20, t+5 \rightarrow t} = \frac{P_{e5+}^{20, t}}{eP_{e5+}^{25, t+5}}$$

$$bTR_{e55}^{20 \rightarrow 15, t+5 \rightarrow t} = \frac{P_{e5+}^{15, t}}{(eP_{e5+}^{25, t+10} * bTR_{e55}^{25 \rightarrow 20, t+5 \rightarrow t})}$$

$$bTR_{e44}^{20 \rightarrow 14, t+5 \rightarrow t} = \frac{P_{e4+}^{15,t}}{eP_{e4+}^{20,t+5}}$$

应用在例子中便可以获得：

表 A1-15 step5 后推转换比率结果

Time	age	bTR_{e22}	bTR_{e33}	bTR_{e44}	bTR_{e55}	bTR_{e66}
$t \rightarrow t-5$	15--19	⊗	⊗	99.1%	26.9%	○
$t \rightarrow t-5$	20--24	○	○	⊗	100.0%	47.4%
$t \rightarrow t-5$	25--29	○	○	○	⊗	75.3%
$t \rightarrow t-5$	30--34	○	○	○	○	⊗

Source: [Springer et al. \(2019\)](#)。

每次迭代都需要重新计算一次 BTR，然后进行升学调整。

（5）获得人口数量的教育分布数据

将获得的 t-5 期的性别-出生队列-教育阶段人口占比信息与 WPP2017 的人口结构信息结合，获得人口数量的教育分布数据。

（6）预测

将一些基年在 2015 年前的国家预测至 2015 年，以产生一致的数据集。这其中包含三个步骤：

一是利用数据进行前推，并考虑生存率调整；

二是利用全球教育趋势情景预测各国 15-34 岁群体的情况。

三是将得到的年龄-性别人口分布进行调整，以精确匹配联合国对年龄-性别人口分布（WPP2017）的估计，同时不改变每个年龄-性别组内的教育分布。

附录二、常见的全球可比的学生认知技能构建研究

（一）文献汇总和全球可比学生认知技能数据库介绍

1. 文献汇总

我们给出了采用文章介绍的基本方法的文献。

表 4-3 关联不同测试项目的不同方法和文献

类别	假设	方法及文献
相同个体		Patel & Sandefur (2020)
相同总体 (国家)	相同总体（总体相同）：合并 测试信息	相同国家（有重复国家）：扩大样本 国家数量
	均值标准差转换 Reardon et al. (2021)	
	基于微 观分布 相似	均值标准差转换 Hanushek & Woessmann (2015); Gust et al. (2024)
		同分位点转换 Sandefur (2018)
		应用于宏观数据的均值标准差转换 Hanushek & Kimko (2000)、Hanushek & Woessmann (2012a)、Hanushek & Woessmann (2012b); Altinok et al. (2018)
	基于测 试项目 差异	线性转换 Angrist et al. (2021) 均值连接 Hanushek & Woessmann (2012a)
		比率连接 Altinok & Murseli (2007); Altinok et al. (2014); Altinok et al. (2018); Patrinos & Angrist (2018); Altinok and Diebolt (2023)
相同题项	题项特征参数等同 Das & Zajonc (2010)、Singh (2014)、Sandefur (2018)	
	题项难度和区别度等同 Sandefur (2018)	

2. 数据库介绍

目前，有不少研究致力于将众多国际性和地区性学生测试得分数据地合并工作，然而只有少数研究公开其研发的数据库，承诺会持续更新的更是少数。我们将已知的有关学生认知技能的国际可比数据库汇总为表 4-4。

首先是世界银行提供的 HLO 数据库，该数据库是目前唯一一个在持续更新的数据库。最新版的 HLO 数据由 Angrist et al. (2021) 搭建，主要使用线性转换方法构建而成，该数据库可直接在世界银行数据银行中搜索获得。如 HLO 数据库之前版本的数据一样，HLO 数据库被应用于世界银行的多份报告之中，并且被世界银行人类发展指数数据库（HCI Database）收录在内。HLO 数据库的

元来源测试项目广泛，使其能够提供了 2000-2017 年 164 个国家和地区按性别-教育阶段（小学、初中）-科目（数学、科学、阅读）划分的国家均值数据，然而由于测试项目限制，该数据是非平衡面板数据，不少国家早年的学生测试得分缺失比较严重。

其次是 [Lim et al. \(2018\)](#) 组建的数据，其先采用线性转换方法构建可比的学生认知技能数据，然后利用时空高斯过程回归（spatiotemporal Gaussian process regression）估计不同国家-性别-年龄群体学习指数（Learning index）。该数据提供了 186 个国家从 1990 到 2016 年的平衡面板数据。然而实际上，拥有测试项目数据的国家只有 132 个，并且也不覆盖所有年份，该数据大部分国家和时间的结果为拟合生成的结果，其可靠性、精确性有待考据，使用该数据需要注意这一点。

然后是 [Altinok and Diebolt \(2023\)](#) 构建的学习技能（Learning Skills）数据。[Altinok and Diebolt \(2023\)](#) 是为了构建质量调整后的受教育年限数据（见下一小节），但在补充材料中同时给出了学生认知技能数据。该数据主要基于比率连接和同分位点连接转换方法，然后同样利用补值程序对缺失数据的年份和国家进行补值。需要注意的是，其虽然在文章中表明其学生认知技能是按性别-教育阶段（小学、初中）-科目（数学、科学、阅读）划分的，但其给出的最终数据为一个值¹³。

最后是 [Gust et al. \(2024\)](#) 组建的数据，主要采用均值标准差转换方法构建而成，是哈努谢克（Hanusheck）团队的最新成果。该数据被附于文章附表之中。由于 [Gust et al. \(2024\)](#) 主要是想构建最新的各国的学生认知技能数据，因此在时间上使用的学生测试项目大多是 2018 或 2019 年的，所以该数据时间要比 HLO 数据库更新一些，只提供各国学生测试得分的单期数据，属于截面数据；在学生群体上，主要以初中为主，没有初中的国家，用小学的进行补充；在科目上，其提供的是数学和科学得分的均值，而没有提供分科目的得分；在国家数目上，其覆盖 159 个国家和地区，在所有数据中也是最少的。

¹³ 并未找到对这一值的介绍，但怀疑是所有情况均值

表 4-4 公开的学生认知技能数据库

	HLO 数据库（新版）	Global Human Capital Estimates	Learning skills	统一尺度下的学生获得数据
方法文献	Angrist et al. (2021)	Lim et al. (2018)	Altinok & Diebolt (2023)	Gust et al. (2024)
转换方法	以线性转换为主，均值标准差转换为辅	线性转换+拟合补值	以比率连接为主，同分位点连接转换为辅，加以补值	以均值标准差转换为主，高拟合补值为辅
元数据来源	TIMSS、PIRLS、PISA、SACMEQ、PASEC、SERCE、TERCE、EGRA	PISA、PIRLS、TIMSS、IAEP、Reading Literacy Study(1991)、SACMEQ、LLECE、PASEC、NAEP、NAS、IQ、MYS	TIMSS、PIRLS、PISA、SACMEQ、PASEC、SERCE、TERCE、EGRA、ASER	TIMSS、PISA、TERCE、SERCE、SACMEQ、PASEC
数据是否分性别	是	否	无区分	否
学段	小学+初中	5-9, 10-14 和 15-19	无区分	以初中为主
科目	数学（Math）、科学（Science）、阅读（Reading）	学习指数（Learning index）	无区分	数学（Math）和科学（Science）的均值
数据跨期与结构	2000-2017 年非平衡面板	1970-2016 年平衡面板	1970-2020 年平衡面板（部分时间为面板）	2018/2019 年横截面
国家和地区数量	164	186	167	159
数据获取	https://datacatalog.worldbank.org/search/dataset/0038001	https://ghdx.healthdata.org/reco/ihme-data/global-human-	https://link.springer.com/article/10.1007/s11698-023-00276-x	Gust et al.（2024）文章附表

capital-estimates-1990-2016

（二）相同个体和相同题项：心理测量学连接（Psychometric Linkage）

据作者所知，Patel & Sandefur（2020）是目前唯一基于相同个体方法所作的研究。在 2016 年，Patel & Sandefur（2020）采用 NEAT（Non-equivalent Groups with Anchor Test）设计（Kolen & Brennan, 2014）将国际上四项最大标准化考试（涉及世界上 80 个国家）的六个不同测试（包括数学测试：TIMSS¹⁴和 PASEC Math、LLECE Math 以及阅读测试：PIRLS 和 PASEC Read、LLECE Read）的部分试题合并为两套混合试卷，并将其分配给印度相同的一组考生。该组学生来自印度的比哈尔邦的六个地区（Bhojpur, East Champaran, Gaya, Jehanabad, Nalanda, and Patna）的 51 所学校，年级横跨 4 到 8，平均年龄为 11 岁，共有 2314 名。

也就是说这一设计结合了相同个体和相同题项。为了更形象地说明这一研究，以数学测试的 TIMSS 和 LLECE Math 为例，我们给出了下图。如图中所示，Patel & Sandefur（2020）从 TIMSS 和 PISA 题库中分别抽取 42 道和 12 道题组成新的试卷，让来自印度的同一批学生做这套混合了 PISA 和 TIMSS 部分题项的试卷。

表 A2-1 混合试卷试题组成

	题库	计算时的题目
TIMSS	60	42
PASEC Math	12	12
LLECE Math	4	4
PIRLS	53	31
PASEC Read	16	16
LLECE Read	4	4

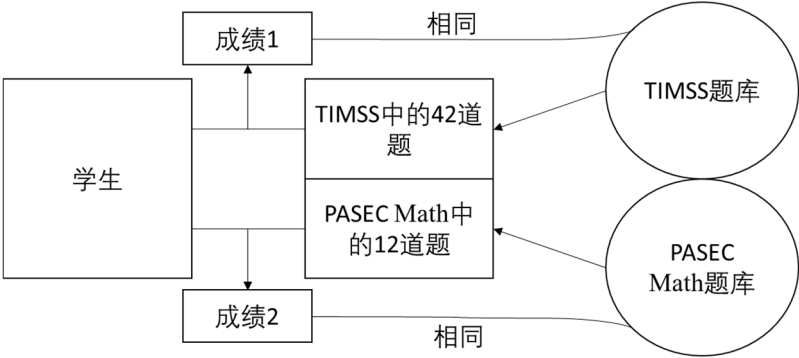
注：计算时的题目删除了少于 30 个人回答正确的 40 道题目¹⁵。

由于混合试卷来自 TIMSS 和 LLECE Math，因此根据相同题项，学生在 TIMSS 部分试题中获得的《成绩 1》将可以与 TIMSS 直接比较；同样《成绩 2》也将和 LLECE Math 的直接可比。又根据相同个体，假设这些学生的能力在一天内是固定不变的，那么《成绩 1》和《成绩 2》将同时衡量学生固定的能力。由此，根据《成绩 1》和《成绩 2》可以通过固定的学生能力构建 TIMSS 和 LLECE Math 的关联，形成 TIMSS 和 LLECE Math 的分数对应表。这一对应表

¹⁴ TIMSS 包括数学和科学，这里使用的是数学。

¹⁵ 原文为 41 道题。

可以用于 TIMSS 和 LLECE Math 之间分数转换。



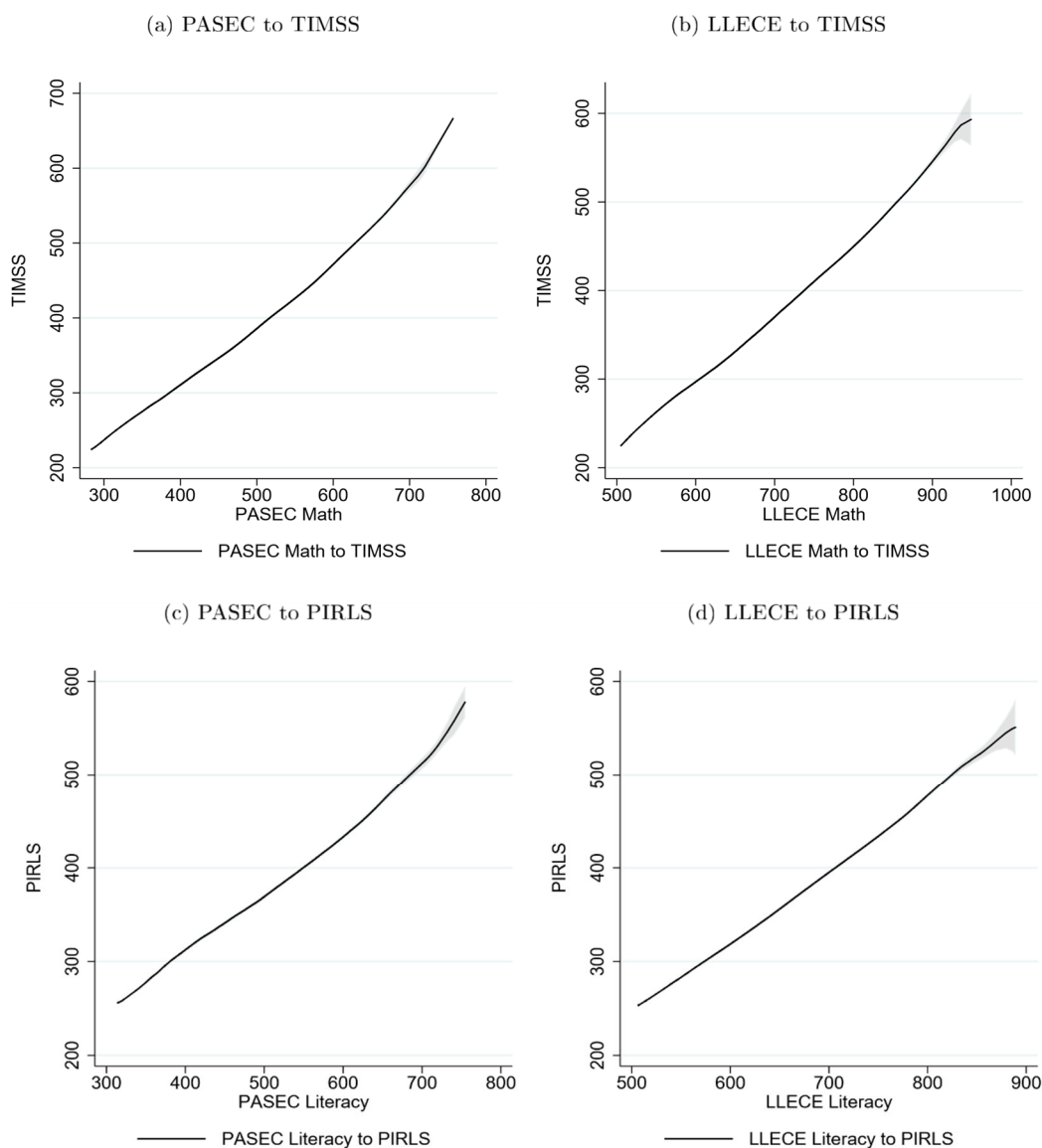
注：作者自绘。

图 A2-1 研究思路示意图

同样，其他也类似，所以文章最终得到各学生的 PASEC Math 和 LLECE Math 与 TIMSS 的得分以及 PASEC Read 和 LLECE Read 与 PIRLS 的得分。

为了两个测试项目的得分对应表，利用学生的得分，使用局部线性回归估计，得到下图的转换函数图形。利用估计获得的局部线性转换函数，便可以将 PASEC Math、LLECE Math 和 PASEC Read、LLECE Read 的得分分别转换为 TIMSS 和 PIRLS 的得分¹⁶。

¹⁶ 需要注意的是，由于 PASEC 主要针对六年级学生，LLECE 主要针对三年级学生，而 TIMSS/PIRLS 主要针对四年级学生，因此即使转换过后也不能直接比较，得分的差异还有可能是由于年级（受教育时间）的差异导致。文中进行了年级调整，以便比较，详情见 Patel & Sandefur（2020）。



Source: [Patel & Sandefur \(2020\)](#)。

图 A2-2 转换函数图形

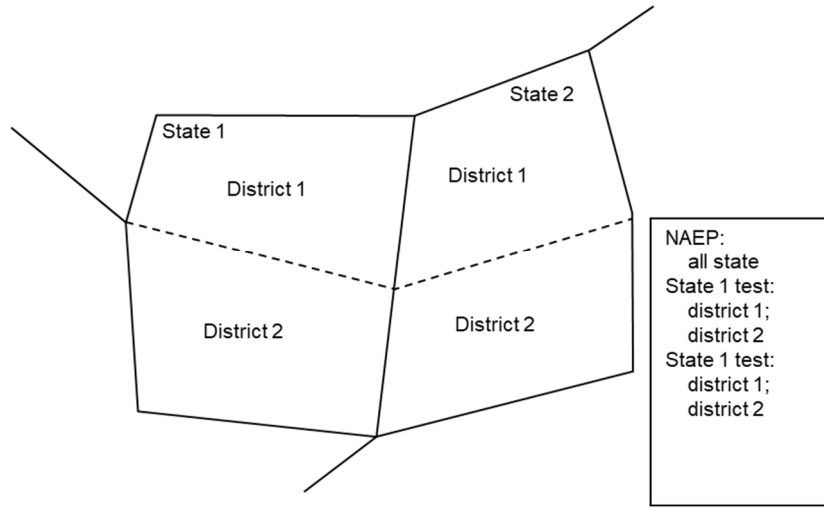
该设计在实施上的优点在于，由于一位同学做多份完整的试题是一个巨大的负担，通过合并多个测试项目的试题形成一份新的测试项目，减轻了学生负担，并且在理论上，该方法也似乎最好的、最有吸引力的方法。然而该研究设计也存在几个较为明显的缺点。一是实施成本太高，实施这一研究设计所需的时间、金钱、精力是巨大的。二是相同题项的数据非常有限。例如，TERCE 只有 4 个选择题，PASEC 数学只有 12 个选择题，而经验法则建议至少需要 30 个选择题才能进行合理的联结 ([Hastedt & Desa, 2015](#))。三是该研究的测试对象为印度的比哈尔邦 (Bihar) 的学生，比哈尔邦 (Bihar) 是印度 ASER 测试项目得分最低的邦之一，而印度学生本身在国际上的表现相对较差，这意味这只有

极低比例的学生能够获得较高的分数，这将损害高分段上转换的质量，乃至可能在较高分段时没有学生分数可供建立分数间的关联（Gust et al., 2024）。

（三）相同总体：整合信息目的下的均值标准差转换

均值标准差转换由 Hanushek 首先使用，但作者认为，在理论上，更符合这一方法设计的研究为 Reardon et al.（2021）。

Reardon et al.（2021）想比较各州之间的学区教育质量，如图中的 *State 1* 的 *district 1* 和 *State 2* 的 *district 2*。然而一方面只有各州自己举办的测试项目能够区分到学区层面，但是由于各州的测试项目完全不同，致使各州间的学区无法直接比较。另一方面，全国层面进行了统一的 NAEP 测试项目，NAEP 只能区分部分州的部分学区，在大多数州都是无法区分到学区。



注：作者自绘。

图 A2-3 研究思路简单示意

此时，若想进行跨州学区教育质量的比较，只需要将由各州组织的考试分数成为转换为全国的 NAEP 尺度上的分数，即将 *State 1 test* 和 *State 2 test* 统一称为 NAEP 的分数。对应的转换公式如下：

$$Score_{cdtlsi}^{neap} = \frac{(Score_{dtlsi}^{state} - \mu_{dtls}^{state})}{\sigma_{dtls}^{state}} \sigma_{ctls}^{neap} + \mu_{ctls}^{neap}$$

其中 c 为州， d 为学区， t 为年份， l 为年级， s 为科目， i 为学生个体。在这已转换过程中，相同总体为各州：每个州都参加了自己的测试和 NAEP 的测试，目的是合并各州测试的学区信息和 NAEP 测试的各州可比信息，实现学区的跨州比较。

另外，虽然上述公式可以转换各年各学科的学生层面的分数，但由于各州的州测试项目和 NAEP 测试项目的内容存在差异，难以保证各州的州测试项目和 NAEP 测试项目分布一致，因此 [Reardon et al. \(2021\)](#) 只关注学区分数的均值，这样可以减少由于方法基本假设不满足带来的影响。在对各州数据进行州内的标准化（ $\sigma_{dtls}^{state}=1$, $\mu_{dtls}^{state}=0$ ）后，所以文章的实际转换公式为：

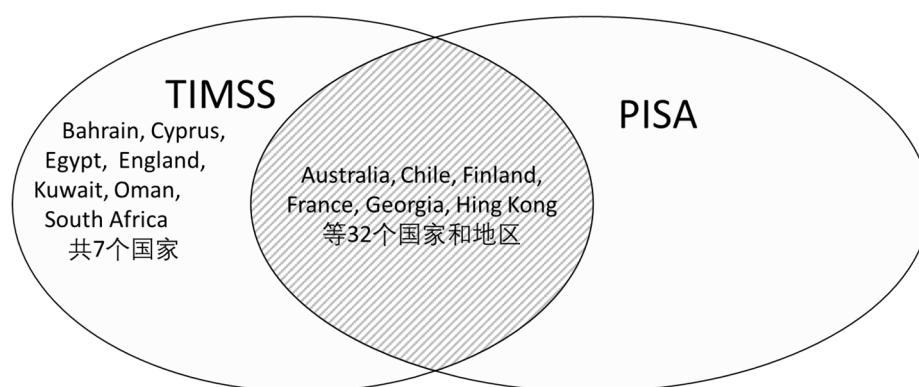
$$\bar{\mu}_{dtls}^{neap} = \bar{\mu}_{dtls}^{state} * \sigma_{ctls}^{neap} + \mu_{ctls}^{neap}$$

最后，[Reardon et al. \(2021\)](#) 也对转换的数据进行了一系列的检验，结果均表明转换后的学区数据拥有较好的质量。

（四）相同国家：扩展样本目的下的均值标准差转换

在国别研究中，最重要的是获得各个国家的学生认知技能数据。这类研究通常以 PISA 作为基准，将 TIMSS 等其他测试的得分转换成 PISA 的得分，以此来获得更多国家的学生认知技能数据。[Gust et al. \(2024\)](#) 就试图将 2019 年的 TIMSS 的数学和科学成绩转换成 2018 年的 PISA 尺度下的成绩¹⁷，以此扩大研究的国家样本¹⁸。

为了该文的思路做直观介绍，先介绍一下 TIMSS2019 和 PISA2018 年的情况。如图中所示，PISA2018 一共有个 79 参与国家和地区；TIMSS2019 包括 4 年级和 8 年级，为了与 PISA 的 15 岁学生对应，使用八年级成绩，一共有个 39 参与国家和地区；这些国家和地区中一共 32 个国家和地区同时参与 PISA2019 和 TIMSS2018。



¹⁷ 与其他地方一致的是，这里默认假设 TIMSS 和 PISA 测量了相同的数学和科学素养。

¹⁸ 根据文中注释，这一方法在 [Hanushek & Woessmann \(2015\)](#) 已被使用，然而由于 [Hanushek & Woessmann \(2015\)](#) 为 OECD 报告，在报告中并没有详细介绍该方法，因此本文以介绍 [Gust et al. \(2024\)](#) 为主。

注：作者自绘。

图 A2-4 PISA2018 和 TIMSS2019 同时参与国家示意图

如在方面部分所述，扩大国家样本目的的研究通常使用同时参与两个测试的国家作为锚点，实现两个测试得分的转换。因此这同时参与了 PISA2018 和 TIMSS2019 的 32 个国家就可以作为转换锚点，形成相同国家，用其所有学生个体的分布来推测 PISA2018 和 TIMSS2019 的学生分布。具体转换公式可写为：

$$p_i = \frac{(t_i - m_{TIMSS}^c)}{s_{TIMSS}^c} s_{PISA}^c + m_{PISA}^c$$

其中 $(m_{TIMSS}^c, s_{TIMSS}^c)$ ； (m_{PISA}^c, s_{PISA}^c) 分别为使用 32 个国家的所有微观学生个体的分数分布估计出来的 TIMSS 和 PISA 的均值和标准差。在获得转换公式后，就可以获得只参与 TIMSS 的国家的学生个体分数所对应 PISA 分数尺度上的学生个体分数。

下图给出了实际转换结果的分布图，图中用作锚点的相同国家的学生分布接近正太分布，这为相同国家的分布推测相同总体的分布提供了依据。同时，可以看出需要转换国家的学生与连接国家的学生分数分布相对位置在转换后基本保持不变，这一点说明转换的有效性。

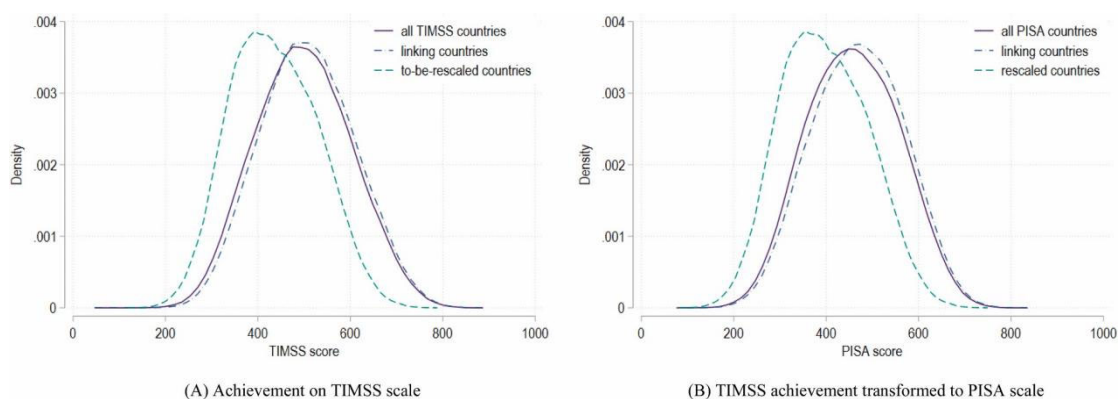
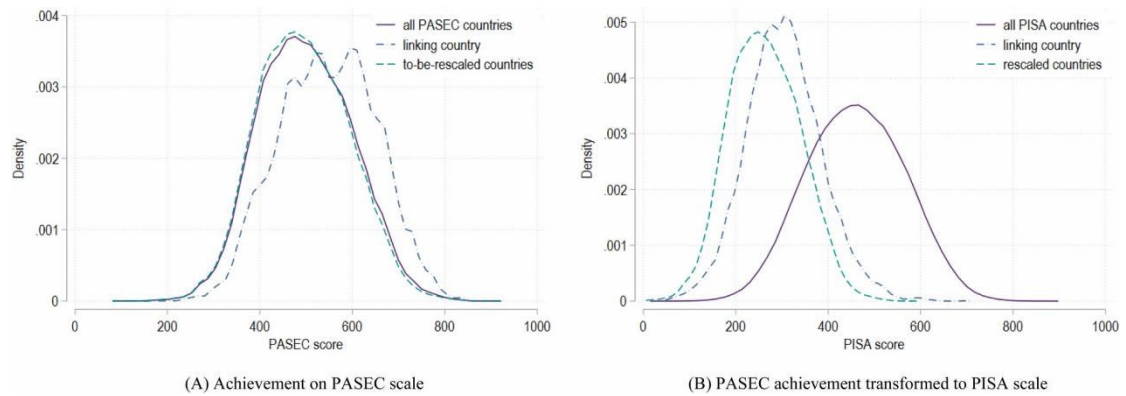


图 A2-5 32 个共同参与国家时转换结果¹⁹

需要注意的是，可供连接的国家数量决定了连接的质量，当连接国家只有一个时，因为测试的差异、国家的特点和学生的抽样，连接可能会产生较大误差。如下图所示，在只有一个共同参与国家时，连接国家在 PISA 和 TIMSS 中的学生得分既不是正太分布，也有很明显的差别，使得转换后，部分学生相对

¹⁹ 由于文章缺乏相应信息，因此不知道图中是数学得分，还是科学得分，还是两者均值。

接连国家的位置有明显的变化。



注：此图中只有一个国家为相同国家：Senegal 同时参加了 PASEC 2014 和 PISA-D。

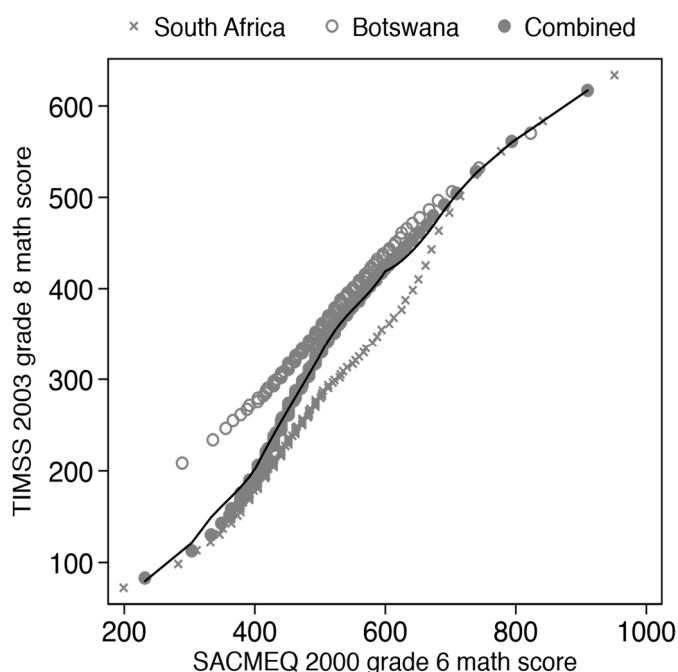
图 A2-6 1 个共同参与国家时转换结果

（五）相同国家：同分位点连接

Sandefur（2018）使用同分位点连接意图将非洲地区性测试项目 SACMEQ II（2000）中 14 个国家和地区的数学分数转换为 TIMSS（2003）尺度下的数学分数²⁰。文章中相同国家为同时参加这两个测试项目的南非（South Africa）和博茨瓦纳（Botswana）²¹。

²⁰ 坦桑尼亚（Tanzania）的桑给巴尔（Zanzibar）单独参加了 SACMEQ II，因此将其排除在外的话，实际上是 13 个国家；Altinok et al.（2018）在构建数据的过程中使用了前平滑同分位点连接，由于文中没有详细方法的具体使用，因此此处只介绍 Sandefur（2018）。

²¹ 使用两个国家来构建连接时，通常外部有效性受到限制，而通常研究需要外部有效性以应用于其他国家来扩充国家样本。



注：其中叉号和空心圆为南非和博茨瓦纳两个国家各自百分位点对应值，实心圆为两个国家合并的学生样本百分位点对应值，实线为线性拟合补值线。

图 A2-7 Sandefur (2018) 研究思路简单示意

由于该文的目的是比较三种方法的结果，因此文章忽略了上面问题。并假设对任何年级或时间的儿童而言，这两个测试项目将产生相同的分数排序²²。在这一假设下，通过对应两个测试项目中两个国家全部学生的分数分布的分位点值，再使用三次样条函数（Cubic Spline Function）进行平滑处理，Sandefur (2018) 得到了用于转换的转换曲线，即上图中的实线。利用这一关系，可以实现从 SACMEQ 得分到 TIMSS 得分的转换。

（六）相同国家：线性转换

基于测试项目差异的线性转换主要由 Angrist et al. (2021) 发展而来。Angrist et al. (2021) 利用此方法更新了 HLO 数据库，构建了 2000-2017 年的全球可比的学习产出数据库。

在实际连接方程参数的估计中，Angrist et al. (2021) 为了使得结果更加精确，通过考虑测试项目科目（subjects: reading、math and science）和教育阶段（schooling levels: primary and secondary）/年级（grade）之间的差别²³，对连接函数进行了进一步改进：

²² 正如方法部分所说，这其实是放松了微观分布相似假设。

²³ 也可以在此基础上考虑性别的差异。

$$\mu_{Y,c,s,l,r} = \alpha_{s,l} + \beta_{s,l}\mu_{X,c,s,l,r} + \varepsilon_{c,s,l,r}$$

其中， c 是同时参与测试项目 X 和 Y 的国家， s 为科目， l 为年级， r 为轮次/时间， $\alpha_{s,l}$ 和 $\beta_{s,l}$ 与国家和测试项目轮次无关的转换参数。如前面所述，2000 以后的测试项目的轮次间（时间上）的可比性，使得可以利用多轮测试项目来估计两个测试项目之间的差异。也因此，随着时间的推移和测试项目的不断开展，将拥有更多可供估计的样本， α 和 β 的估计也将更精确。

在具体实施过程中，共分为三步：第一步是将不同时间的不同考试进行匹配，形成不同轮次的考试对 (X_r, Y_r) ，如 2018 年 PISA 和 2019 年的 TIMSS 形成的 $(PISA_{2018}, TIMSS_{2019})$ 。在操作中，如果测试项目的时间间隔在五年内，就可以被认为是同一轮测试项目。但在轮次匹配时，应尽量使得测试项目的时间间隔最小。在 [Angrist et al. \(2021\)](#) 的所有情况中，最大的时间窗口期都在四年内。

第二步是利用同一轮次中同时参与两个测试项目的国家的样本（即相同国家）估计关联函数，如第一轮次的美国、葡萄牙等，第二轮次的美国、匈牙利等。第三步为利用估计获得的 $\alpha_{s,l}$ 和 $\beta_{s,l}$ ，将只参与测试项目 X 的国家某年级某学科的分转换成目标测试项目 Y 的分数。该方法除了第一步繁琐外，其他步骤实施起来较为简单。

（七）相同国家：比率连接

基于测试项目差异的比率连接主要来自世界银行的一系列文献（[Altinok & Murseli, 2007](#)；[Altinok et al., 2014](#)；[Altinok et al., 2018](#)；[Patrinis & Angrist, 2018](#)）。这一方法在世界银行的一系列文献被用于构建 2000（20 世纪 90 年代）年前的可比数据以及 2000 年的可比数据。

2000（20 世纪 90 年代）年前的数据主要是时间上不可比，需要找到一个时间上可比的锚点，作为相同国家。美国的 NAEP 因其时间上的可比，通常被用来作为这个。[Altinok et al. \(2014\)](#) 以例子的形式直观的介绍如何以 NAEP 为相同国家，使用比率连接横向和纵向构建可比数据。

表 A2-2 以一个国家为锚点的比率连接

Country	TIMSS 1995	NAEP	Adjustment	TIMSS 1995 adjusted
Australia	509		509* (548/492)	567

Canada	521		521* (548/492)	580
Japan	581		581* (548/492)	647
Norway	498		498* (548/492)	555
United States	492	548	492* (548/492)	548

如，上表。因为只有一个国家（美国）作为锚点，因此比率为美国在其他测试上（比如 TIMSS1995）的成绩和 NAEP 成绩的比值（即 548/492）。其他测试以这种方式转换为 NAEP 的分数，实现了各测试之间的可比，同时由于 NAEP 本身在时间上是可比的，因此最终得到的数据实现了横向和纵向的可比。

在该系列文献的最新一篇 [Patrinós & Angrist \(2018\)](#) 中，为了构建 2000-2017 年的数据。2000 年后，各测试以共同题项实现了时间上的可比，因此只需要保证横向国家上的可比。同时，也因为时间上的可比，也保证了各测试之间的差异在各轮次间保持不变，也就是说转换系数与轮次无关，可以利用多年数据来估计转换系数。

该文首先统一两个测试项目的分数尺度，将两个测试项目下各年份的分数都转换为均值为 500 标准差为 100 的尺度下的分数。这一步是 [Patrinós & Angrist \(2018\)](#) 独有的，在 [Altinok et al. \(2018\)](#) 以及之前的文献中都还未曾出现过。

其次，将两个测试项目下各年份进行的测试项目划分为某一轮次 r ，并利用下式，计算出相同轮次、相同科目和相同被测学生的层次（Level）中同时参与两个测试项目的相同国家的各国分数的均值。

$$\mu(X_{rsl}) = \frac{1}{n_{rsl}} \sum_{c \in X_{rsl} \cap Y_{rsl}}^r X_{crsl}$$

$$\mu(Y_{rsl}) = \frac{1}{n_{rsl}} \sum_{c \in X_{rsl} \cap Y_{rsl}}^r Y_{crsl}$$

接下来，在获得以上要素后，利用下式获得转换系数：

$$d_{sl} = \frac{1}{r} \sum_{r=1}^r \frac{\mu(X_{rsl})}{\mu(Y_{rsl})}$$

最终，利用转换系数，对不同科目、不同时间和不同被测试项目学生年级的分数进行转换：

$$Y_{cysl} = \frac{X_{cysl}}{d_{sl}}$$

该方法是早期 HLO 数据库构建的主要方法，直到最近，HLO 的构建方法才变为线性转换。

（八）相同题项：题项特征参数等同

Das & Zajonc (2010) 使用项目反应理论连接方法将印度两个州的数学测试项目转换成 TIMSS 尺度下的数学分数。该研究基于的是 2005 年在印度拉贾斯坦邦 (Rajasthan) 和奥里萨邦 (Orissa) 的 288 所学校内开展的 9 年级学生数学测试项目。该数学测试项目由 36 道题目组成，全部来自 TIMSS1999 年八年级的数学测试项目。在利用该数据计算学生数学素养时，36 道题目的特征参数完全来自于 TIMSS1999，因此，使用 36 道题所获得的两个州的分数可以与 TIMSS 中各国的学生数学素养直接进行比较。

开展于 2002 年的 Young Lives Study 调查在四个国家（埃塞俄比亚、印度(安德拉普拉德什州)、秘鲁和越南）的两个出生队列中进行，并在 2006 年和 2009 年进行了两次追踪调查。在两次追踪调查中，添加了认知技能测试项目，其中在 2006 年对 12 岁学生的调查中加入 10 道数学题项，其中有 6 道题来源于 TIMSS。利用从 TIMSS 中获得的这 6 道题项的特征参数，Singh (2014) 将其余四道题的特征参数和学生数学素养当作未知参数，估计出了与 TIMSS 可比的 2006 年 12 岁学生的数学素养得分。

相比较而言，采用同样方法的 Sandefur (2018) 的研究则复杂一些。Sandefur (2018) 意图将非洲地区性测试项目 SACMEQ II (2000) 中 14 个国家的学生数学素养得分转换为 TIMSS (2003) 尺度下的学生数学素养得分。实现这一目的的基础是 SACMEQ 2000 的学生测试项目与 TIMSS 有足够多的相同题项。然而如下图所示，SACMEQ 2000 年学生测试项目与 TIMSS1995 年 7 和 8 年级测试项目仅有 3 个相同题项，但 SACMEQ 2000 的教师测试项目分别与这两个测试项目的相同题项都较多，因此，Sandefur (2018) 利用 SACMEQ 2000 的教师测试项目作为中间测试项目，将 SACMEQ 2000 的学生数学素养得分转换为 TIMSS (2003) 尺度下的学生数学素养得分。

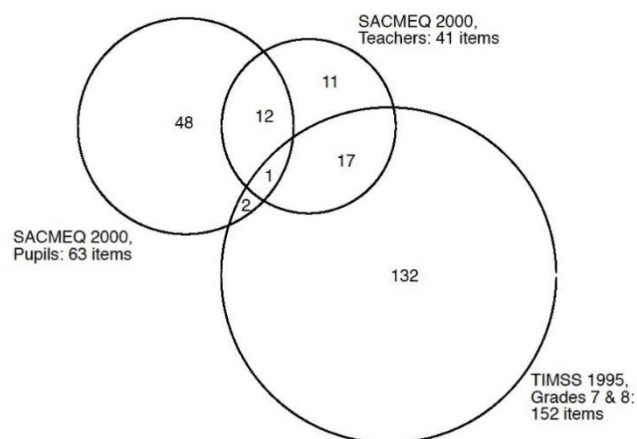


图 A2-8 TIMSS1995 和 SACMEQ 重叠测试题项示意图

具体而言，其分为两步，使用 SACMEQ 教师中与 TIMSS1995 的重叠题项的数据，利用公式估计出教师的数学素养，其中相同题项的特征参数与 TIMSS1995 保持一致；第二步，使用 SACMEQ 学生与教师的相同题项数据，利用公式估计出学生的数学素养，其中相同题项的特征参数与 SACMEQ 教师保持一致。需要注意的是，该研究上述程序只利用了相同题项，并没有如前面方法部分介绍的，利用相同题项重新估计非相同题项的特征参数，在此基础上再计算学生的数学素养。

（九）相同题项：题项难度和区别度等同

Sandefur (2018) 年使用项目反应理论连接方法，使用题项难度和区别度等同将非洲地区性测试项目 SACMEQ II (2000) 中 14 个国家的数学分数转换为 TIMSS (2003) 尺度下的数学分数。若是只是简单的两个测试之间的转换的话，其具体实施和方法介绍部分一致。但图 3-11 中所述的复杂的相同题项情况使得这一方法在文中稍显复杂。

其第一步是根据均值标准差 (Mean-sigma) 法分别参考测试与中间测试，中间测试和目标测试之间的 A_{rl} 和 A_{lt} ， B_{rl} 和 B_{lt} 。

其次利用下式获得最终的 A_{rt} 和 B_{rt} ：

$$A_{rt} = A_{rl}A_{lt}$$

$$B_{rt} = A_{lt} * B_{rl} + B_{lt}$$

如前面所述，Sandefur (2018) 只利用了相同题项来估计潜变量，并没有利用相同题项重新估计非相同题项的特征参数，在此基础上再计算学生的数学

素养。因此，该研究可以直接使用公式（20） $\theta_t = A_{rt} * \theta_r + B_{rt}$ 直接转换学生的分数。

（十）相同国家：应用于时间维度的 Hanushek 转换方法发展概述

之前已有所提及，基于分布的认知技能可比构建由 Hanushek 的一系列文章发展而来，主要包括 Hanushek & Kimko (2000)、Hanushek & Woessmann (2012a)、Hanushek & Woessmann (2012b)。由于在 20 世纪 90 年代之后，项目反应理论才被应用于各测试项目中，使得跨期可比，若使用在此之前的测试项目，需要考虑时间上的可比性：即需要将不同年份的测试项目可比化²⁴。

Hanushek & Kimko (2000)是该系列的第一篇文章，在这篇文章中，方法包含三个步骤：第一步，通过比率转换将各年份的、各年龄（或年级）、各科目的分数进行转换，使得均值等于 50，即：

$$\begin{aligned} Meanscore_{c,l,s,t}^{first} &= \beta_{l,s,t} * Meanscore_{c,l,s,t} \\ s.t.: \mu(Meanscore_{c,l,s,t}^{first}) &= 50 \end{aligned}$$

其中 $Meanscore_{c,l,s,t}$ 为原始的国家 c ，年级 l ，在 t 时的学科 s 的总成绩； $Meanscore_{c,l,s,t}^{first}$ 为经过一次处理后的成绩。据文章介绍，这一步依赖于一个强有力的假设：世界科学和数学成绩的跨期均值是不变的，参加测试的国家是从世界分布中随机抽取的。

第二步，将美国作为相同国家，使用美国在时间上可比的 NEAP²⁵分数，采用均值连接，使得：

$$Meanscore_{c,l,s,t}^{second} = Meanscore_{c,l,s,t}^{first} - Meanscore_{US,l,s,t}^{first} + NEAP_{l,s,t}$$

$Meanscore_{c,l,s,t}^{second}$ 的经过第二次处理后的成绩。这一步其实对所有时期 t 的各国各学科成绩进行上下平移，使得美国的得分等于其当年的 NEAP 的得分。这一步的意图是利用美国 NEAP 得分的变化，调整各国成绩以实现各国成绩在

²⁴ 之前的测试项目较少，所以时间上的可比性是研究的主要目标。

²⁵ 事实上，2000 年前的测试面临两个问题，一是时间上的可比性，二是不同测试之间的可比性。为了实现时间上的可比性，这类研究需要一个时间上可比的锚点，由于美国长时期的开展 NAEP，并且其设计上可以在时间上比较，因此这类研究通常将美国 NAEP 分数作为时间的锚点。

时间上的可比。

最后，利用加权平均计算各国的教育质量指标，权重为各国标准误的逆（the (normalized) inverse of the country-specific standard error (σ)):

$$SQ_c = \frac{1}{\sigma_c} \sum_t \sum_l \sum_s Meanscore_{c,l,s,t}^{second}$$

这一步是将所有年份所有学科和所有年级的分数算一个加权平均值，用以代理各国较长时期的教育系统质量。

该文章的处理较为粗糙，从以下几点中可见：其一，文章只涉及均值的可比，没有对标准差进行调整，也就是说只是对难度进行了调整，没考对区分度进行调整。其二，在假设各年份的测试项目形成的分布是一样的（[Altinok et al., 2018](#)），只是均值不一样，这一假设在 2000 年前是很难满足的。其三，在第一步时，采用比率转换的方式使得各年份的均值等于 50，这一做法的合理性也值得商榷。最后，该研究在第二步时使用的是均值连接，并且相同国家只有一个国家：美国，因此，转换的质量直接依赖于美国得分在时间上的随机性误差的大小。

[Hanushek & Woessmann \(2012a\)](#) 在 [Hanushek & Kimko \(2000\)](#) 的基础上，进一步对方法进行了改进，主要的改进就是在利用标准差进一步对区分度加以考虑。其方法主要分为四步：

第一步是获得在时间上可比的基准。时间上的可比基准基于的是 NAEP 数据。其本身就是在时间上可比的，只不过文中是以 PISA2000 为基础进行可比，因此需要将 NAEP 的时间可比数据转换成 PISA2000 尺度下的时间可比数据。由于这里所说的可比，通常是均值可比，因此，这一问题就变成将 NAEP 各年的均值差距转成 PISA2000 尺度下的均值差距。

为了解决这一问题，假设存在一个 NAEP 分布，其形状不变，每年变化的只是其水平位置，即 $SD^{US,NAEP}$ 不变， $NAEP_t^{US}$ 。 $SD^{US,NAEP}$ 未知，但可以使用每年的 $SD_t^{US,NAEP}$ 在时间上的均值 $\overline{SD}^{US,NAEP}$ 进行估计。在此假设下，1999 年的 NAEP 分布为 $(\overline{SD}^{US,NAEP}, NAEP_{1999}^{US})$ 。若将其他年份的均值数据放置于 1999 年的分布上，那么此时上述问题便可以看作将 1999 年²⁶分布上的一些分数转换为

²⁶ 选择 NAEP1999 是因为这是可获得的，距离 PISA2000 最为接近的数据。但这也假定了 1999 到 2000 年间美国的学生认知技能水平无变化。

PISA2000 年的分数。

此时，相同总体为美国，两个测试项目为 PISA2000 和 NAEP1999，根据基于微观分布相似的均值标准差转换，考虑年级 l^{27} 学科 s 公式可以写为：

$$D_{lst}^{US} = O_{s,2000}^{US,PISA} + \frac{NAEP_{lst}^{US} - NAEP_{ls1999}^{US}}{\overline{SD}_{ls}^{US,NAEP}} SD_{s,2000}^{US,PISA}$$

由此，各年数据到 1999 年的差距为：

$$T_{lst}^{US} = D_{lst}^{US} - O_{s,2000}^{US,PISA}$$

其中 $O_{s,2000}^{US,PISA}$ 为美国在 PISA2000 中科目 s 的原始分数， $NAEP_{lst}^{US}$ 为不同年级 1、不同科目 s 、不同时间 t 的 NAEP 分数； $\overline{SD}_{ls}^{US,NAEP}$ 为年级 1、科目 s 的 NAEP 分数标准差在一定时期内的平均值； $SD_{s,2000}^{US,PISA}$ 为 PISA2000 年科目 s 的标准差。 D_{lst}^{US} 为经过均值标准差转化后美国 NAEP 各年的分数。 T_{lst}^{US} 为 PISA2000 尺度下，各年 NAEP 分数到 1999 年 NAEP 的差距。由于 PISA 只有 15 岁（对应 8 年级）的学生成绩，因此，这里隐含假设了不同年级的学生的得分是不变的。经过以上处理，可以获得类似下图的一个分数基准：

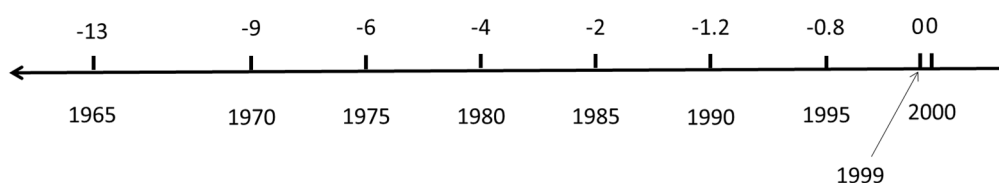


图 A2-9

第二步是忽略时间维度的变化，获得横向上的可比结果。以教育系统稳定的 13 个 OECD 国家（OSG）作为相同国家，利用均值标准差转换，将其他国家的测试项目分数转换成 PISA2000 分数（教育系统稳定是指这 13 个国家虽然在各种测试项目中的得分不一致，但在各种测试项目中的相对位置都是一致的）²⁸。

²⁷ 文中为年龄组 a 。

²⁸ 需要注意的是，这 13 个国家并不一定都出现在每个测试中。

$$C_{lst}^c = O_{lst}^{OSG,PISA} + \frac{O_{lst}^c - O_{lst}^{OSG}}{SD_{lst}^{OSG}} SD_{s,2000}^{OSG,PISA}$$

其中 O_{lst}^c 表示国家 c 在 t 期测试项目中年龄组 a 科目 s 的原始分数； O_{ast}^{OSG} ， SD_{ast}^{OSG} 为各中测试项目中 OSG 国家均值的均值和标准差； $O_{ast}^{OSG,PISA}$ 和 $SD_{s,2000}^{OSG,PISA}$ 分别为 PISA2000 中 OSG 国家均值的均值和标准差。这一步是将均值标准差转换应用于国家层面³⁰。这一转换如下图所示。

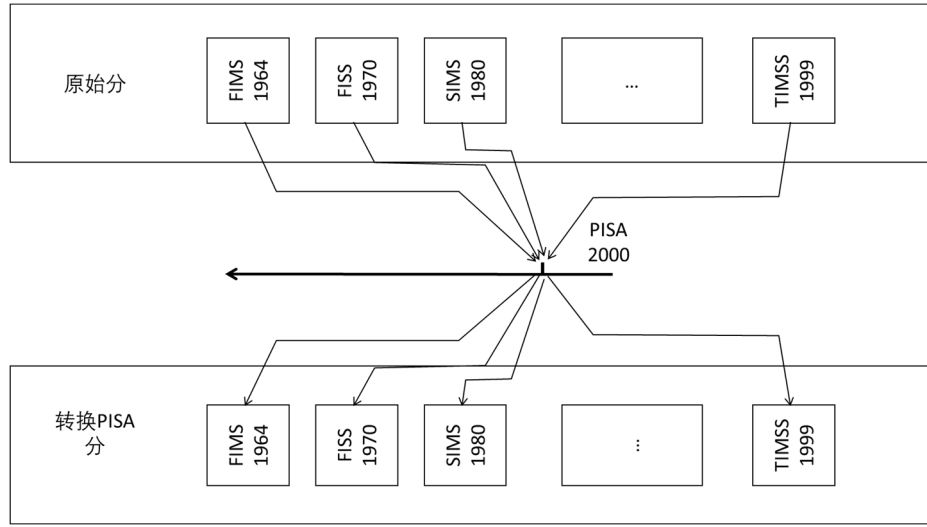


图 A2-10

经过这一步转换，我们获得了“忽略时间上的变化下的”横向上的可比结果，譬如 TIMSS1999 和 PISA2000。所有转换的都是 PISA2000 尺度下的分数，都成为 2000 年也意味着时间上变化信息的缺失，这正是“忽略时间上的变化”这一前提条件的来源。当我们的数据只有一个年份时，此时我们就已经获得了想要的结果；但因为数据属于不同的年份，还需要进一步处理。

第三步是结合纵向可比的标准和横向可比的数据相结合，从而构建在时间上和在国家上都可比的数据。这一部主要是将横向可比的各轮数据以美国为基准进行调整：

$$I_{lst}^c = C_{lst}^c - C_{lst}^{US} + T_{lst}^{US}$$

²⁹ 在 Altinok et al. (2018) 中，这一方法被称为线性转换。

³⁰ 在文章中使用的是 OSG 国家中各个国家均值的均值和标准差（单位是国家），但从分布的角度讲，应当使用 OSG 国家所有学生的均值和标准差（单位是学生个体）。因为 OSG 各国的学生数目不一致，这两个的结果会略有不同。因此，实际上这一步可以视作为将各国作为参与测试的个体，然后对这些国家个体形成了两个测试得分分布进行转换。

其中 I_{lst}^c 为最终的结果。 C_{lst}^c 为忽略时间维度的变化下的横向可比结果， C_{lst}^{US} 为其中美国的结果； T_{lst}^{US} 为时间上的基准。在调整方式上，这一步和 [Hanushek & Kimko \(2000\)](#)中的第二步十分相似。这一步骤也可以用下图直观展示。

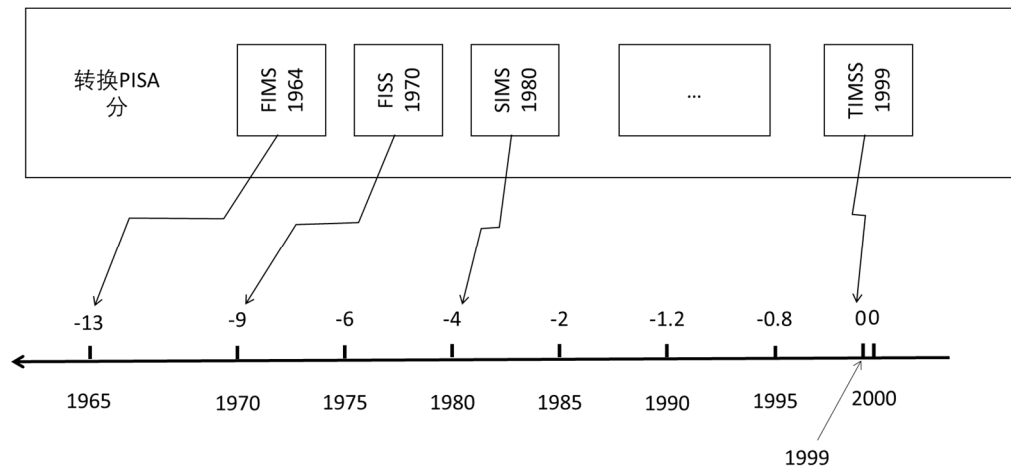


图 A2-11

最后，同样是在不同年份-年级-科目上取平均，获得一个整体的教育系统质量衡量指标。

应用这一方法，[Hanushek & Woessmann \(2012a\)](#)对 1965-2003 年间 12 个国际性测试项目分数进行可比转换，然后通过所有年份、所有年级、所有学科的平均构建了 77 个国家的学生认知技能数据³¹。

其后，[Hanushek & Woessmann \(2012b\)](#)进一步加入了拉丁美洲的地区性测试项目（6 个国家），使得样本扩充至 86 个国家。

在 1997 年，9 个国家参与了 LLECE（对 4 年级的测试）；在 2006 年，13 个国家参与了 SERCE（对 6 年级的测试），其中有 6 个早先参与了 LLECE。因此一共有 16 个拉丁美洲国家参与了地区性测试，其中有 7 个国家又参与了国际性测试项目。为了将地区性测试的得分转换为国际性测试的得分，如下图所示，[Hanushek & Woessmann \(2012b\)](#)进行了如下几个步骤：

³¹ 文章在构建方法时，在第一步时强调了使用 1999 年及之前的 NAEP 数据构造时间上可比的基准，而对于 2000 及以后已经实现了时间上可比的 PISA、TIMSS 和 PIRIS 数据并没有说明数据处理方法，可能是根据第二步公式进行改编。

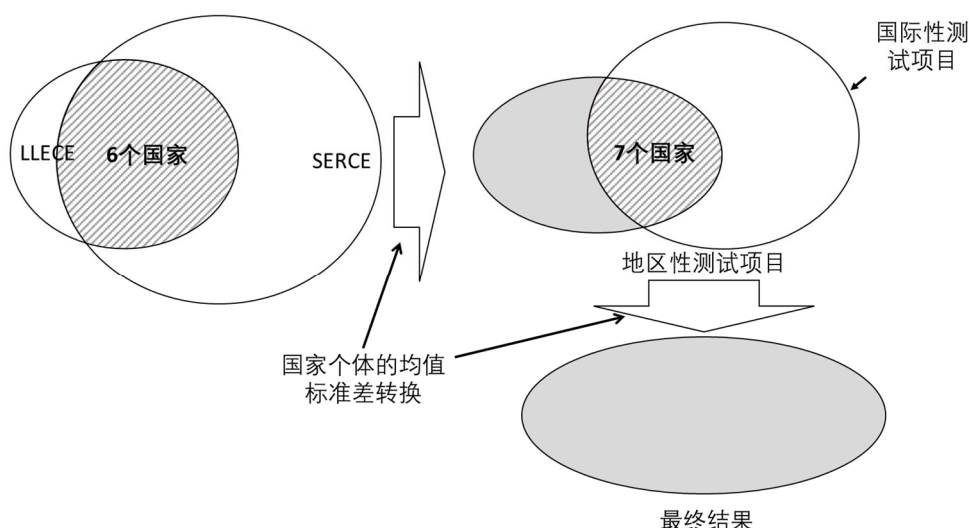


图 A2-12

第一步是为了整合两个地区性测试项目（LLECE 和 SERCE）。首先根据同时参加两个地区性测试项目的 6 个国家得分均值的均值和标准差，将两个测试项目的分数进行标准化转换，要求标准化后两个测试项目中的 6 个国家均值得分的均值为 0，标准差为 1。然后对这标准化后的测试均值按国家分别求平均，获得一个汇总的地区性测试项目得分。简单地理解这一步，可以假设存在另一个只有这 6 个国家参与的测试项目，6 个国家的均值得分的均值为 0，标准差为 1。然后将根据相同国家和均值标准化转化方法，分别将 LLECE 和 SERCE 的得分转换为这个假设的测试项目的得分。

第二步将同时参与国际性和地区性测试项目的 7 个国家设置为相同国家，使用利用均值标准差转换方法将地区性测试项目分数转换成国际性地区分布³²。

最后，文章使用转换后的地区性测试项目分数，而非国际性测试项目的分数，作为 7 个共同参与国家的分数。文章作者认为这可以解决国际性测试项目可能对低表现水平国家衡量不准确、低表现水平国家内部相对位置不可靠的问题³³。

通过改良后的方法，[Hanushek & Woessmann \(2012a\)](#)和 [Hanushek & Woessmann \(2012b\)](#)完善了国家个体的、基于微观分布相似的均值标准差转换方

³² 这里的七个国家的国际性测试项目的结果是 [Hanushek & Woessmann \(2012a\)](#)中最终构建的数据结果。

³³ 作者利用文章中提供的数据重现了文章的数据构建过程，经过这两个程序后自己计算的数据和提供的结果依然有细微的差别，但两者的 R2 已经得到 0.99。

法，构建了国家认知技能衡量指标，被应用于经济增长的分析之中，得到了一些重要的发现。但文章的一些操作其实影响了文章的科学性。其一，[Hanushek & Woessmann \(2012b\)](#) 的第一步隐含的假设是：4 年级参与的在 1997 年实施的 LLECE 和 6 年级参与的 2006 年实施的 SERCE 有相同的得分的分布，这是一个较强的假设。其二，地区性测试使用的是 1997 年和 2006 年的测试项目，而国际性测试使用的是 [Hanushek & Woessmann \(2012a\)](#) 的结果，为 1965 年至 2003 年的测试项目，测试项目存在时间的不一致。其三，[Hanushek & Woessmann \(2012b\)](#) 和 [Hanushek & Woessmann \(2012a\)](#) 一样，都是在事后的学生认知技能数据对之前（1960-2000 年）的经济数据作回归分析，在理论上就存在不合理性。

附录三、数量和质量结合：不同类型的认知技能生产函数的文献

（一）文献汇总和数据库介绍

1. 文献汇总

下表给出了使用以上基本方法的文献。

表 5-1 角度-函数-文献对应表

投入产出	函数形式	文献
投入端	对数 ($\beta * \ln(nq)$)	Kaarsen (2014)
	线性 ($\beta * nq$)	Hanushek & Zhang (2009)
产出端	乘积 (nq)	Filmer et al. (2020)
		Angrist et al. (2020)
		Glawe & Wagner (2022)
		Altinok & Diebolt (2023) ³⁴

2. 数据库介绍

目前，可利用的 QAYS 国际数据库有两个，一个由 [Filmer et al. \(2020\)](#) 开发构造，另一个是由 [Altinok & Diebolt \(2023\)](#) 构造。由于这两个数据库在构造过程中，都使用了学生认知技能，而学生认知技能被认为是“Learning”结果。因此，这两个数据库又将经质量调整后的受教育年限称为“经学习调整后的受教育年限（Learning-adjusted Years of Schooling, LAYS）”。

由 [Filmer et al. \(2020\)](#) 开发的 LAYS 数据被世界银行采纳，其同样被收录于人力资本指数数据库中，用于人力资本指数的计算。世界银行 HCI 数据库提供

³⁴ 该文提供的技术细节较少，并且存在矛盾点，读者应当注意。

2010、2017、2018、2020 年四年的 174 个国家的非平衡面板数据。需要注意的是 Filmer et al. (2020) 利用的是受教育年限数据为 25-29 岁人口的，接近于其学生认知技能数据年份，因此该数据衡量的是一国当前的教育系统³⁵。

另一个是由 Glawe & Wagner (2022) 构造，其与 Filmer et al. (2020) 不同之处主要在于其构建了 33 个国家 1995-2015 年（5 年间隔）的平衡面板数据。

第三个数据为 Altinok & Diebolt (2023) 构建的数据，其提供了 120 个国家从 1970 至 2020 年（5 年间隔）的平衡面板数据。与前两者相比，Altinok & Diebolt (2023) 其有两个不同点。其一，其受教育年限使用的是 25-65 岁人口的，因此该数据指向的是一国人资本存量，而非该国当前的教育系统。其二，该数据进行了较多的拟合估计，如 2020 年各国平均受教育年限、年份和国家部分缺失的学生认知技能，从这里可以看出，其数据并不纯粹，可能存在较大的误差。由于其是新构建的数据库，因而数据被应用和接受程度尚且不知。

表 5-2 教育质量或调整后的平均受教育年限数据库

方法文献	LAYS 数据库		
	Filmer et al. (2020)	Glawe & Wagner (2022)	Altinok & Diebolt (2023)
方法（函数形式）	公式（5-11）	公式（5-11）	公式（5-11）
元数据来源	文章中：TIMSS 或 PISA		TIMSS、PIRLS、PISA、
	学生认知技能	TIMSS 或 PISA	SACMEQ、PASEC、LLECE、SERCE、TERCE、EGRA、ASER
	后续更新：Global Dataset on Education Quality（2020 Update）	Barro-Lee 数据+ CSL 数据库 ³⁶ （25-29 岁）	Barro-Lee 数据（25-64 岁）
受教育年限	Barro-Lee 数据（25-29 岁）		
时间跨度	文章中：2015 后续更新：2010-2020 年非平衡面板	1995-2015 平衡面板（5 年间隔）	1970-2020 年平衡面板（5 年间隔-部分时间为面板）
覆盖国家和地区数目	文章中： 后续更新：174	33	120
数据获取	https://datacatalog.worldbank.org/search/dataset/0038030/Human-Capital-Index	未知	https://link.springer.com/article/10.1007/s11698-023-00276-x#Sec110

³⁵ 在数据的后续更新中，说明文档并未介绍一些细节，因此尚且不知其为何能够使得覆盖国家达到 174 个。

³⁶ <https://www.parisschoolofeconomics.eu/en/cohendaniel/international-educational-attainment-database/>

（二）投入端：对数型认知技能生产函数

在考虑科目和时间的情况下，Kaarsen（2014）在教育质量在各年级不变的假设下，认为教育质量在短期内也不随时间变化而变化：今年三年级的教育质量和明年的教育质量是一样的，这一假定容许使用不同时间的数据来估计教育系统的质量。另外，作者认为技能生产在学科间表现不同：对科学素养的培养和对数学素养的培养是不一样的，因此需要认知技能生成函数中的参数应随学科而变化。

文章的一个关键是决定函数的形式。这需要对认知技能与年级的变化有充分的了解。文章通过两个说明验证认知技能和年级之前的关系。

其一，文章利用 TIMSS 数据中 3 年级和 4 年级，7 年级和 8 年级的数据分别计算了小学和初中每增加一年学生成绩的变化，如下表所示，初中的一年的变化增加的认知技能要小于小学的³⁷，这表明学校教育对认知技能的边际产出是递减的：随着年级的增加，相同教育质量所能产生的认知技能提高是减少的。

表 A3-1

	Math primary	Math secondary	Science primary	Science secondary
Increase	66.56	29.40	59.99	36.46
St. dev.	49.80	45.84	44.87	33.88
Increase/st. dev.	1.34	0.64	1.34	1.08

其二，文章认为，得分水平与分数的变化可以提供一些信息：如果在分数较高的水平，年级间分数的变化越小，这就表明随着投入的增多，其能够带来的增值是减小的。如下图所示，在刨除图三图四中特异值（主要是南非和哥伦比亚）的影响，所有的图中的拟合线都将是向下倾斜的。

³⁷ 文章认为小学和初中的得分是不可比的，为此分别除以标准差，对分数标准化后的结果进行比较，其结果依然一致。

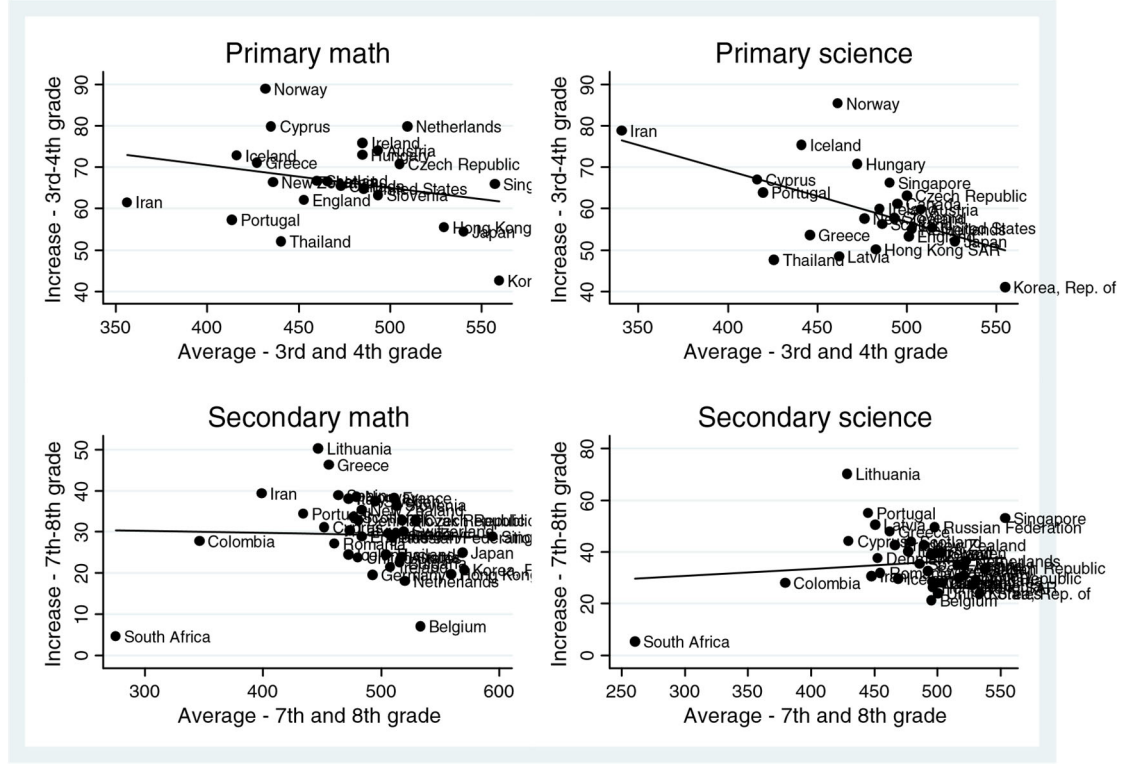


图 A3-1

经过上述两项验证，可以确认学校教育对认知技能的边际产出是递减的：随着年级的增加，相同教育质量所能产生的认知技能提高是减少的。在这些假定下，作者认为学生认知技能生产函数是对数形式，可以写为：

$$CS_{t,c,l,s} = \alpha_s + \gamma_s * \ln(n_l * q_c)$$

式中 $CS_{t,c,l,s}$ 为在 t 时国家 c 的年级 l 的学生在学科 s 上的得分，式中 q_c 只随国家的不同而不同，学科间的不同有 α_s 和 γ_s 进行捕捉。为了对教育质量进行估计，对上式进行改写：

$$CS_{t,c,l,s} = \alpha_s + \gamma_s * [\ln(n_l) + \ln(q_c)] = \alpha_s + \gamma_s * \left[\ln(n_l) + \sum_{j=1}^{81} D_{j,c} * \ln(q_c) \right]$$

上式是关于 α_s 、 γ_s 和 q_c 的函数。使用非线性最小二乘法进行估计，可获得各国的教育质量（ q_c ）。需要注意的是，只有存在不同年级 l （如三年级和四年级）的分数时才能对上式进行估计，因为只有这时才能估计出 γ_s ，继而估计出 q_c 。

在获得各国的教育质量，并且将美国设置为参照国家后即可构造质量调整系数：

$$Q_c^b = \frac{q_c}{q_{us}}$$

使用 TIMSS 的 1995、1999、2003、2007 和 2011 年三四年级和七八年级的数学和科学分数，[Kaarsen \(2014\)](#) 获得了质量调整系数，如图所示。[Kaarsen \(2014\)](#) 发现各国的教育质量存在较大差异：美国一年的教育（Schooling）成果相当于很多低收入国家三年及以上的教育（Schooling）。

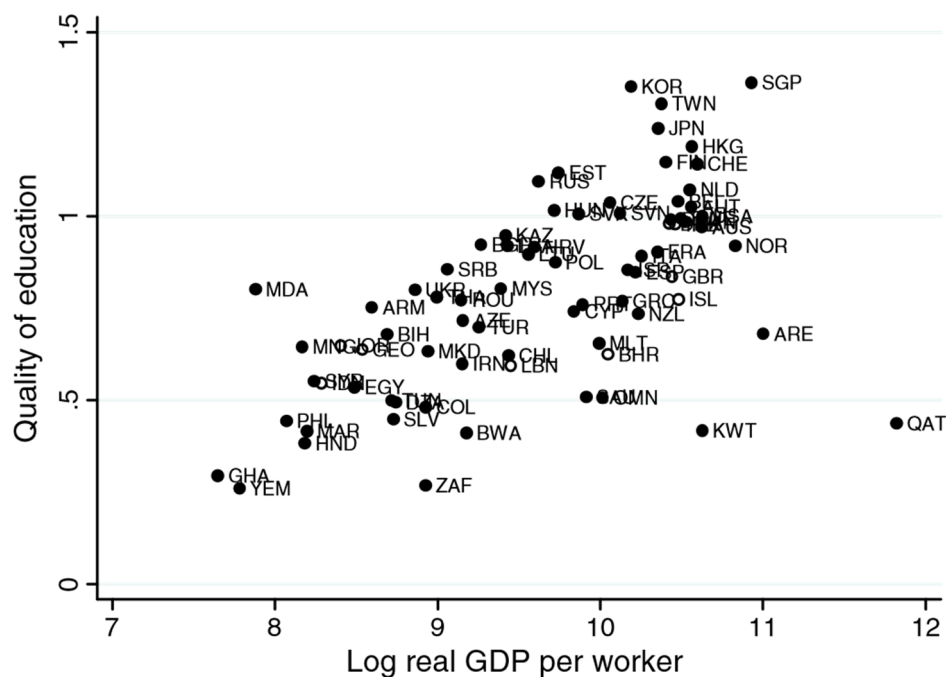


图 A3-2 研究思路简单示意

由于这篇文章的主要目的不是为了构建质量调整后的平均受教育年限（Quality-adjusted Years of Schooling），而是利用质量调整后的平均受教育年限构建人力资本存量，因此文章只给出了人力资本存量结果，而非调整后的平均受教育年限。人力资本存量根据以下公式计算：

$$H = e^{\theta(n,q)}, \theta(n,q) = \frac{\theta}{\eta} (nq)^\eta$$

其中 nq 为质量调整后的平均受教育年限。

下图为文章的最终结果：

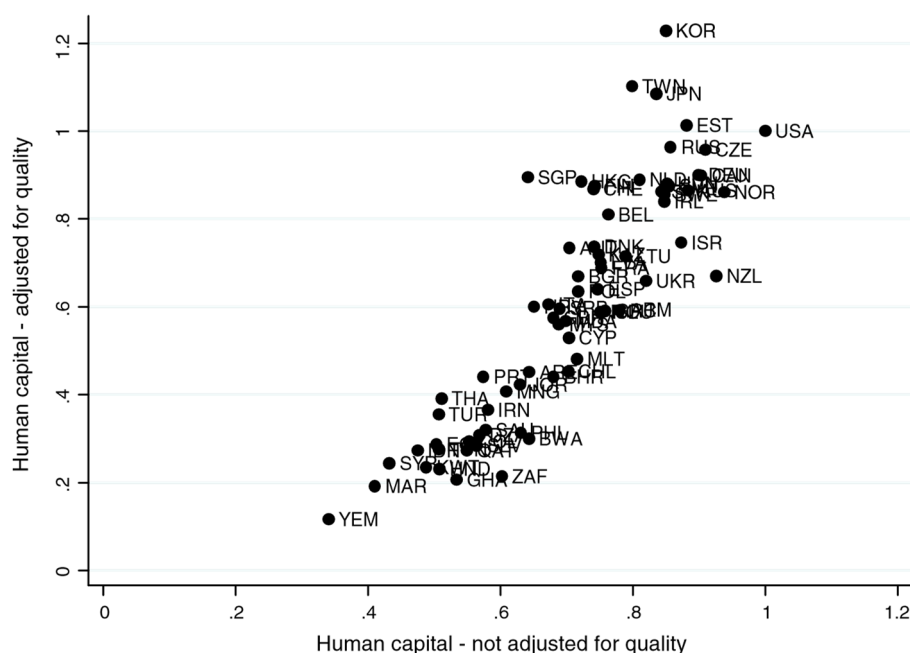


图 A3-2

利用人力资本存量，文章对世界经济增长差异进行了分解，结果显示，在对人力资本进行质量调整后，人力资本对世界经济差异的贡献提升了 22%。

这篇文章的问题在于，其使用现在的教育系统质量，调整该国的平均受教育年限，用于探索人力资本和经济增长的关系：由于质量是根据 1995-2011 年的 TIMSS 学生素养得分数据计算而来的，而平均受教育年限是国家所有年龄人口的均值，因此除非教育质量在各国长期保持不变，否则这一调整本身就会因用现在的教育质量调整之前的教育数量而存在误差。

（三）产出端：累积型认知技能生产函数：Filmer et al.（2020）

Filmer et al.（2020）³⁸认为教育质量（ $q_{c,s}$ ）为一年学校教育可获得的分数增值，某一年级参与的国际性测试项目的分数为之前所有时期学校教育带来的增值的累积。并且教育质量在不同年级是不变的，这也就是说一年级可获得的认知技能增值和三年级的是一样的³⁹，在这一假设下，忽略了其他因素的影响后，认知技能生产函数边可以采用简单累积形式⁴⁰：

³⁸ 该方法也被 Angrist et al.（2020）进一步扩展，用于微观的各类干预实验/政策评估的成效统一化比较之中。

³⁹ 这一假设暗指每年增加的认知技能都是一样的，事实上，正如 Filmer et al.（2020）文中所述，这一假设只有 5 年级至 10 年级时较为符合。

⁴⁰ 文中使用的都是单一年份的数据，也就是说文章并未考虑时间，因此时间下标被去掉。

$$CS_{c,a,s,l} = n_{c,a,l} * q_{c,a,s}$$

其中， $CS_{c,a,s,l}$ 为某个国家 c 的人口队列 a 在某一年级 l 时科目 s 的分数， $n_{c,a,l}$ 为该人口队列年级所对应的教育数量， $q_{c,a,s}$ 为这一队列人口所接受的教育质量。由于统一测试项目经常测量的是同一年级或年龄的人口，因此上式中的下表 a 可以省略。与 Kaarsen (2014) 中“教育质量在科目间保持相同，但认知技能生产不同”不一样的是 Filmer et al. (2020) 认为不同科目的教育质量也不相同。

由此：

$$q_{c,a,s} = CS_{c,a,s,l}/n_{c,a,l}$$

利用此式，可以获得各国-科目的教育质量。在获得教育质量后，可以构建教育质量调整系数；在文中新加坡被选作对照国家。然后结合出生队列 a 的平均受教育年限： $S_{c,a}$ ，就可以获得质量调整后的平均受教育年限（Learning-Adjusted Years of Schooling, LAYS）。

下图中的 LAYS 是作者使用 TIMSS2015 年的数学得分和 Berrro-lee Dataset 中 25-29 岁人口的平均受教育年限⁴¹计算而得到的调整后的平均受教育年限，与未调整的平均受教育年限相比，大部分国家的取值都有所下降，并且对于像南非等教育质量较弱的地方，其取值下降最大。

⁴¹ 文章对 25-29 岁年龄人口的平均受教育年限的年份一致没有提及，作者怀疑是 2010 年或 2015 年。

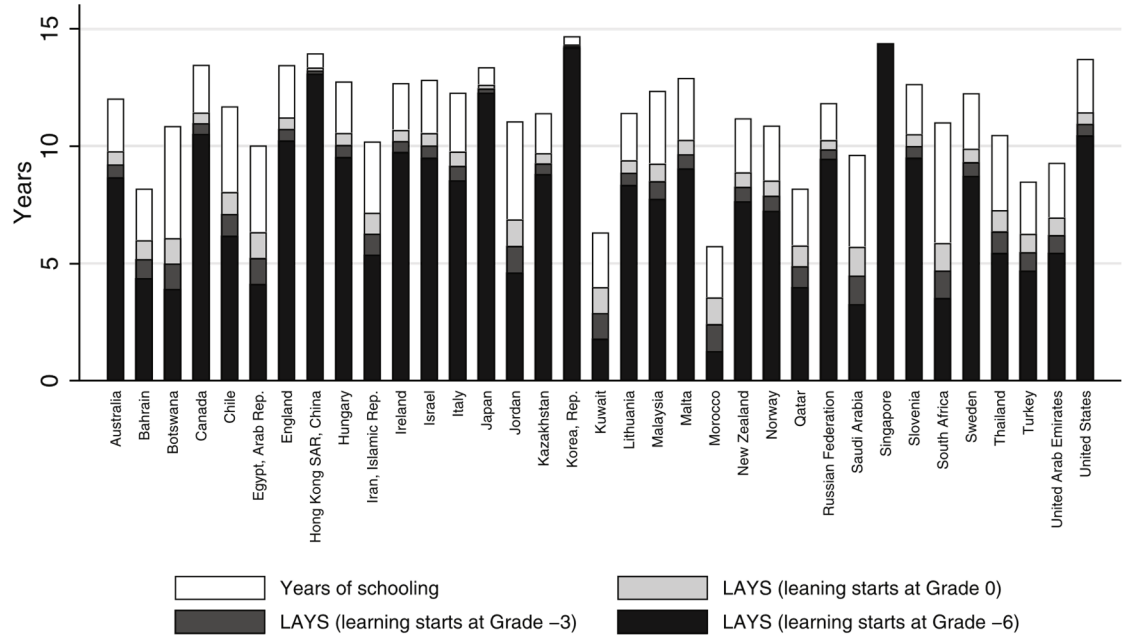


图 A3-4 研究思路简单示意

该文也注意到忽略其他因素的影响可能会对结果造成不好的影响。对此，文章从另一个角度对这进行了回应。其认为学生并不是从入学时开始学习技能，即使是在3岁时，一些孩子已经开始获得最基本的技能（Filmer et al., 2020）⁴²。这就出现了学习时间（Years of Learning）和学校教育时间（Years of Schooling）的差别。当考虑不同的开始学习的时间时，由此，公式可以改编为：

$$CS_{c,l,s} = (n_{c,l} + q_a) * q_{c,s}$$

此时， $q_{c,s} = CS_{c,l,s} / (n_{c,l} + q_a)$ 。

由于文中， q_a 不随国家而变化， Q_c^b 的结果不会发生变化，需要对教育质量调整公式进行调整：

$$LAYS_c = S_c * Q_c^b - [q_a * (1 - Q_c^b)]$$

从上图的结果中可以看出，随着假设受教育年龄的不同，教育所贡献的平均受教育年限在不同国家有不同程度的减少⁴³。

需要注意的是，与 Kaarsen（2014）不同的是，Filmer et al.（2020）的平均受教育年限选择的是25-29岁人口的平均受教育年限、学生认知技能数据使用的

⁴² 这相当于考虑了学前教育，然而家庭教育等其他因素依然在文章的考虑范围内。

⁴³ 论文对这一结果的稳健性做了大量检验，包括更换数据、科目、其他参照国家等等，详情见 Filmer et al.（2020）。

TIMSS2015 数据。因此，如果说 Kaarsen (2014) 是对一个国家的人力资本存量进行调整的话，那么 Filmer et al. (2020) 更多地是对一个国家的人力资本增量进行调整。实际上，Filmer et al. (2020) 在文中不断地强调其衡量的是教育系统 (Education System)，而没有使用人力资本一词。此外，虽然 25-29 岁人口与 TIMSS2015 所衡量的 2015 年教育系统质量还是存在时间上的错位，但相对于整个劳动力人口数据而已，这一时间错位无疑小了很多。

然而文章依然有一些值得讨论的地方。其一，文章用学习时间替代入学时间，这相当于考虑了学前教育，然而家庭教育等其他因素依然在文章的考虑范围内。其二，当获得教育质量后，通过公式 (31) 获得 25-29 岁人口的成人认知技能，直接用成人认知技能来衡量该国这一出生队列的人力资本，应当比调整平均受教育年限更有吸引力。

(四) 投入端：线性型认知技能生产函数：Hanushek & Zhang (2009)

在使用线性认知技能生产函数的 Hanushek & Zhang (2009) 中，研究关注的重点与前两个研究有明显的不同，该研究利用的是有 13 个参与国家的国际成人素养调查 (International Adult Literacy Survey, IALS) 数据，IALS 项目的很多被测者出生在第二次世界大战前后，教育质量在这前后发生了巨大变化，因此 Hanushek & Zhang (2009) 想通过认知技能生产函数估计各个出生队列的教育质量，分析其教育质量的变化，并利用教育质量对个体的受教育年限做调整，其称为质量等同的受教育年限 (Quality-equivalent Schooling, 简称 QES)。

与前两个相比，教育质量和对受教育年限调整都是文章的关注点；其次，文章在构建认知技能生产函数时考虑了其他因素对认知技能的影响；然后，文章使用的是成人认知技能数据，而非学生数据，并且其使用的是个体微观数据，而非国家总体均值数据；最后相较于之前认为教育质量在 (短) 时间上是不变的，Hanushek & Zhang (2009) 强调教育质量的变化性。

教育质量的变化应该体现在出生队列 (16-25, 26-35, 36-45, 46-55 以及 56-65) 上。据此，文章的认知技能生产函数写为：

$$CS_{i,c,a} = \beta_c \sum_{a=1}^5 n_{i,c,a} * q_{c,a} + X_{ic}\gamma_c + \varepsilon_{ica} = \sum_{a=1}^5 n_{i,c,a} * \beta_{c,a} + X_{ic}\gamma_c + \varepsilon_{ica}$$

$$\beta_{c,a} = \beta_c * q_{c,a}$$

式中 $CS_{i,c,a}$ 国家 c 中出生队列 a 的个体 i 的认知技能得分， $n_{i,c,a}$ 为该个体的受教育年限， $q_{c,a}$ 为出生队列的教育质量， X_{ic} 为其他影响因素，在本文中，主

要包括教育系统选择性、健康和技能折旧。

需要注意的是，上述公式只能够估计出 $\beta_{c,a}$ ，不能估计出 $q_{c,a}$ ，这一问题对文章来说并不大，因为文章的目的不是估计出一个绝对的教育质量，其可以利用某一参照出生队列进行对比而消除掉 β_c 。文章中的这一参照出生队列选为 26-35 岁，即 $a = 2$ ，质量调整系数即为：

$$Q_{c,2}^{c,b} = \frac{\beta_{c,a}}{\beta_{c,2}} = \frac{q_{c,a}}{q_{c,2}}$$

在获得了质量调整系数后，作者对所有个体的平均受教育年限进行了调整：

$$QES_{i,c,b} = S_{i,c,b} S Q_{c,2}^{c,b}$$

Hanushek & Zhang (2009) 文章与之前两篇文章的最大不同之处是通过加入相关控制变量来控制其他投入对认知技能产出的影响。文章主要包括以下四点：其一，随着各国教育的发展，教育系统的选择性逐渐减少。其通过加入 $(1-w_a)$ （个体所进入教育阶段的入学率）来控制这种选择性的逐渐减小，当同出生队列的入学率越大时，其选择性越小；其二，其通过加入母亲受教育程度和同出生队列的婴儿死亡率（IMR）来控制家庭投入，婴儿死亡率（IMR）可以捕捉个体早期发展中的平均健康状况对获得认知技能（Learning）的影响。其三，其通过加入年龄的多次项来控制由于成长和老去所带来的心理和身体变化等导致的技能获得和丧失的影响。其四，其认为在职培养或干中学同样会产生影响，但因为测试项目更多地关注非职业的技能，因而其假设工作经验对测试项目分数的影响是有限的，省略工作场所学习不会影响学校教育对能力素养贡献的估计。

表 A3-3

Age Cohort	Chile	Czech Republic	Denmark	Finland	Germany	Hungary	Italy	Netherlands	Norway	Poland	Sweden	Switzerland	United States
16–25	.169 [.008]**	.165 [.014]**	.153 [.010]**	.119 [.008]**	.073 [.012]**	.159 [.013]**	.142 [.011]**	.067 [.008]**	.191 [.013]**	.136 [.012]**	.092 [.010]**	.103 [.009]**	.130 [.011]**
26–35	.159 [.007]**	.146 [.013]**	.131 [.008]**	.094 [.006]**	.071 [.010]**	.138 [.013]**	.143 [.011]**	.055 [.007]**	.177 [.011]**	.125 [.010]**	.089 [.009]**	.090 [.008]**	.134 [.010]**
36–45	.154 [.007]**	.131 [.012]**	.129 [.008]**	.090 [.006]**	.060 [.011]**	.125 [.012]**	.145 [.011]**	.047 [.007]**	.174 [.011]**	.134 [.009]**	.069 [.008]**	.078 [.008]**	.138 [.010]**
46–55	.162 [.008]**	.133 [.013]**	.118 [.008]**	.097 [.007]**	.056 [.011]**	.110 [.011]**	.155 [.012]**	.042 [.008]**	.168 [.012]**	.135 [.009]**	.069 [.008]**	.079 [.009]**	.142 [.009]**
56–65	.162 [.009]**	.128 [.015]**	.119 [.010]**	.089 [.009]**	.034 [.014]*	.090 [.013]**	.148 [.013]**	.021 [.009]*	.141 [.013]**	.137 [.010]**	.053 [.010]**	.074 [.011]**	.143 [.010]**

我们复制了原文中教育质量的回归结果，从中可以看出，对于样本中的几乎所有国家，学校教育对素养技能的贡献在更近期的出生队列中更高。这一结果表明了教育质量的存在随时间的变化，并且主要表现为教育质量提高。并且，质量调整后的受教育年限，[Hanushek & Zhang \(2009\)](#) 估计了教育收益率。

附录四、成人认知技能构建文献

（一）文献汇总和成人认知技能数据库介绍

1. 文献汇总

如前所述，目前文献只有一篇。

表 6-1 文献汇总表

方法	文献
出生队列匹配与关联	Égert et al. (2024)

2. 数据库介绍

[Égert et al. \(2024\)](#) 一文构建了 2020 年的 17 个国家 15-64 岁人口 54 个国家 15-39 岁人口的人力资本存量数据，其中 15-39 岁人口数据覆盖的国家样本数要比 PIAAC 多 18 个国家。受学生认知技能数据的局限，其构建的人力资本存量指标更多的还是发达经济的，只包含较少的中等收入经济体。

另外，此论文构建的数据没有公开，需要自行向作者索取。

（二）出生队列匹配法：Égert et al. (2024)

现有研究的局限常常是数据限制，为了扩展现有成人认知技能数据，[Égert et al. \(2024\)](#) 根据学生和劳动力市场上成人的关系，以丰富的学生认知技能数据，来扩展成人认知技能数据。

在简单了解其方法思想后，我们来介绍文章具体实施的一些过程和细节。首先，虽然成人认知技能数据只有 PIAAC 数据，但学生认知技能数据其既使用了 PISA，也使用了 Global Dataset on Education Quality。这使得学生认知技能数据可以延伸至 1965 年，因此有足够多的学生来对应成人数据的不同出生队列，

尤其是年龄较大的出生队列。

文章方法的第一步是通过历史的学生数据构建某一年份新的出生队列，其公式为：

$$STS'_{a,t} = STS_{t-[(a-1)*5-e]} \text{ for } a = 1, 2, 3, \dots, 10$$

其中 STS 为原始的学生测试成绩， STS' 对应后的学生测试成绩； t 为对应后的年份（如方法部分的 2017 年），这一年份应当是该国拥有 PIAAC 数据的年份； a 为出生队列，其定义与之前一样； e 为年份的调整，取值为 1 或 2，通常，学生测试项目实施的间隔为 3 年或 4 年，因此不可能正好与 PIAAC 数据年份相差 5 年，该调整目的是使得经过计算后的年份为学生测试项目实行的年份。比如，PIAAC2014 年的 60-65 年龄组将和 1965 年的学生成绩相对应。

如下图所示，对应的学生认知技能得分和成人认知技能得分是相关的，这给予了文章进行进一步操作的信心。

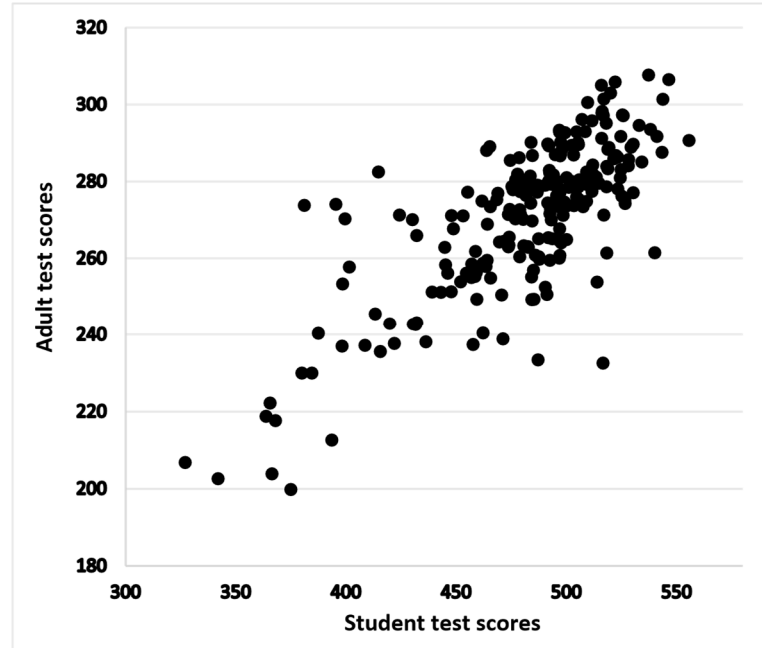


图 A4-1

第二步是构建关联函数，文章的关联函数设定如下：

$$\begin{aligned} \log(ATS_{c,a,t}) = & \alpha + \beta \cdot \log(STS'_{c,a,t}) + \sigma \cdot \log(STS'_{c,a=8,9,t}) + \theta \cdot \log(STS'_{c,a=10,t}) \\ & + \lambda \cdot \log(MYS_{c,a=1,2,\dots,10,t}) + \varepsilon_{c,t} \end{aligned}$$

其中 $ATS_{c,a,t}$ 为对应国家 c 在 t 时出生队列 a 的成人认知技能数据； $STS'_{c,a,t}$ 对

应国家 c 在 t 时出生队列 a 的学生认知技能对应值， MYS 为平均受教育年限。
Égert et al. (2024) 通过将出生队列拆分成三个分布，以此考虑随着年龄的增加可能存在的技能磨损折旧问题。此外，其认为成人认知技能由成人受教育的数量和质量共同决定，因此在模型中加入了平均受教育年限作为教育的数量，而学生认知技能作为教育质量⁴⁴。

下表是文章的估计结果，从表中各模型的拟合优度来看，模型能够很好的拟合学生和成人认知技能的关系，尤其是模型（5），其 R^2 达到了 0.934。

表 A4-1

		(1)	(2)	(3)	(4)	(5)
Dependent variable: log(adult test scores (ATS))		log(ATS) = f(log(STS'), log(MYS))				
α	Constant	2.992*** (0.35)	1.596*** (0.29)	4.214*** (0.23)	1.509*** (0.26)	3.732*** (0.25)
β	log (STS'), all cohorts (baseline effect)	0.438*** (0.06)	0.650*** (0.05)	0.234*** (0.04)	0.603*** (0.04)	0.278*** (0.04)
δ	log (STS'), cohorts 50-59 (additional effect)		-0.006*** (0.00)	-0.010*** (0.00)	-0.006*** (0.00)	-0.009*** (0.00)
θ	log (STS'), cohorts 60-65(additional effect)		-0.013*** (0.00)	-0.016*** (0.00)	-0.012*** (0.00)	-0.015*** (0.00)
λ	log (Mean years of schooling (MYS))				0.152*** (0.02)	0.083*** (0.01)
Adjusted R-squared		0.808	0.671	0.922	0.727	0.934
Number of observations		222	222	222	220	220
Number of countries		34	34	34	34	34
Country fixed effects		YES	NO	YES	NO	YES

利用模型（5）的估计系数，可以通过下式获得某个国家某一时间点（文中为 2020 年）的成人认知技能平均值⁴⁵，也即人力资本存量：

$$\begin{aligned}
 \log(hc_{16-65y,c}) &= \alpha + \sum_{a=1}^7 \frac{wpop_{a,c}}{wpop_c} \cdot \beta \cdot \log(STS'_{c,a,t}) \\
 &+ \sum_{a=8}^9 \frac{wpop_{a,c}}{wpop_c} \cdot (\beta + \delta) \cdot \log(STS'_{c,a,t}) + \frac{wpop_{10,c}}{wpop_c} \cdot (\beta + \theta) \\
 &\cdot \log(STS'_{c,10,t}) + \lambda \cdot \log(MYS_c)
 \end{aligned}$$

其中 $\frac{wpop_{a,c}}{wpop_c}$ 为出生队列人口占总人口的比重⁴⁶。

需要说明的是，虽然理论上可以使用这一方法构造所有出生队列的数据，

⁴⁴ 有关认知技能生产函数的讨论见下一节。

⁴⁵ 需要注意的是，式中用了模型（5）除固定效应外的其他估计参数。因为必须有这个国家的学生和成人认知技能数据（PIAAC）才能估计该国的固定效应，因此对于需要通过回归补值的国家而言，是无法获得其固定效应大小的。

⁴⁶ 文中虽然未曾提及，但这一信息页因来自联合国 WPP。

但由于学生数据的限制，最终只有 17 个国家（其中 15 个 OECD 国家）能够拥有 15-65 岁各出生队列的成人认知技能数据。但如果将出生队列缩小至 15-39 岁时，样本国家数目可以达到了 54 个。

下图给出了构建的数据和原始 PIAAC 成人认知技能数之间的关系，可见，不论是那个年龄分组，都具有相当大的相关性。

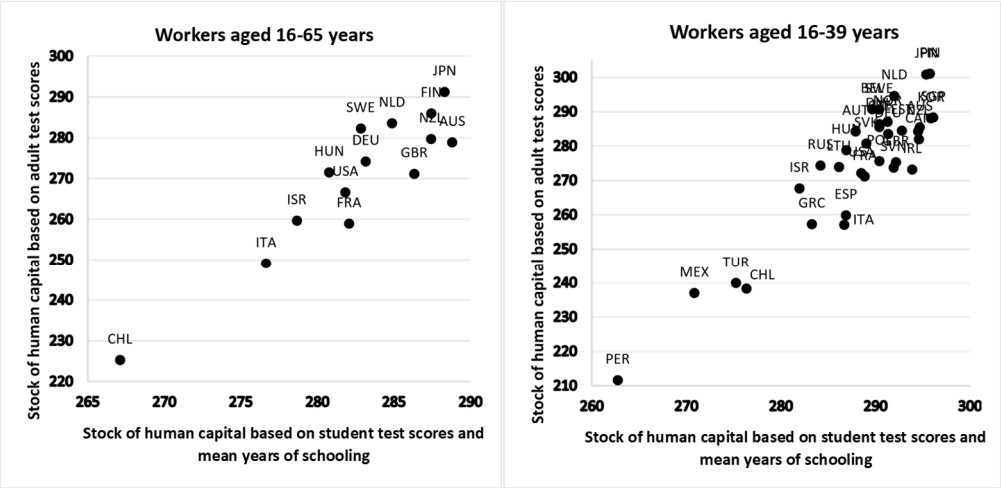


图 A4-2

下图给出了其结果数据的大致情况：

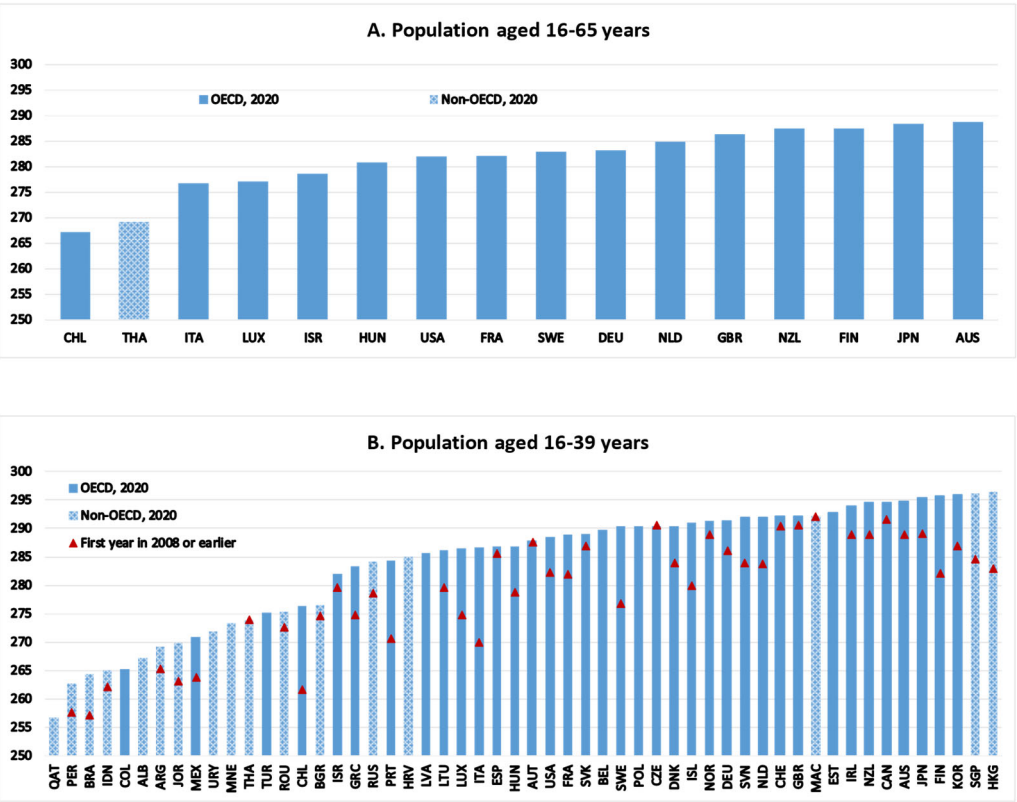


图 A4-3

Source: [Égert et al. \(2024\)](#)。

[Égert et al. \(2024\)](#) 开创性地通过历史上参与 PISA 的学生的人力资本增量精确地获得了 2020 年时各国的人力资本存量，但这一方法由于数据的限制，23 个 OECD 国家（占比 60%）无法获得 55-65 岁成年人的认知技能数据，不过这一问题无疑会随着 PISA 等国际性测试项目的继续实施而得以解决。

此外，除了方法部分所说的局限外，该文章认为 PISA 学生的认知技能能够当作人力资本的流量，并没有在关联方程中考虑高中教育、高等教育对人的认知技能的发展作用⁴⁷；并且在处理过程中，对年份时间的处理太多粗糙，我们无法相信利用疫情前的数据获得的转换关系，可以用来估算疫情之后年份的成人认知技能。

附录五、劳动力市场上的工资信息文献

（一）文献汇总和数据库介绍

1. 文献汇总

该部分的文献汇总如下：

表 7-2 文献汇总

方法	文献
微观路径下的明瑟收入方程	Schoellman (2012)
宏观路径下的人力资本价格	Martellini et al. (2024)

2. 数据库介绍

本文所关注的两篇研究都没有将自行构造的数据作为成果公开，只有 [Martellini et al. \(2024\)](#) 在数据附件中提供了文章图表的复现数据，因而此处无法展开介绍。

⁴⁷ 加入国家固定效应，确实可以对这些情况加以控制，但由于是使用样本中部分拥有 PIAAC 数据的过来估计连接方程，因此其余国家无法获得其对应的固定效应，因此文章最后虽然使用的是加入固定效应后的模型的估计系数，但在构建数据过程中并没有加入国家固定效应。

（二）微观路径下的明瑟收入方程：以教育收益率衡量各国教育系统质量

Schoellman（2012）估计来自在美国不同国家的移民来估计不同国家的教育收益率，以此衡量各国的教育质量。

文章构建的明瑟收入方程如下：

$$\log(W_{US}^{c,i}) = \gamma_{US}^c + \mu_{US}^c * S_{US}^{c,i} + \beta * X_{US}^{c,i} + \varepsilon_{US}^{c,i}$$

其中上标*i*为个体，*c*为出生国家，下标 US 表示为信息是在美国被观测。*W* 为平均小时工资， μ_{US}^c 不同国家的教育收益率， $S_{US}^{c,i}$ 为个体受教育年限， $X_{US}^{c,i}$ 为其其他一些控制变量。

在样本选择上，Schoellman（2012）使用 2000 美国 5% 人口普查数据。该人口普查数据包含来自 130 个国家的移民⁴⁸，每个移民来源国家都至少有 10000 个移民个体。人口普查中含有一个衡量教育获得的指标，按照惯例将其重新编码为受教育年限。不过人口普查没有直接提供在哪里获得学校教育的信息，作者通过使用年龄、移民年份和受教育程度的信息来推断哪些移民是在移民国家完成的学业，并使用在预期完成学业后 6 年才进行移民的样本，以保证这些移民都是在来源国家完成的学业。此外，只选择 18-65 岁受雇佣的、在过去一年至少工作 30 周，每周工作 30 小时的工人，以确保其能工资能够反映市场对其的出价。最终样本包括 4100 万美国人和 21 万移民。与此同时，数据还将英语水平、移民年份、工作经验及其平方项和普查居住地区、性别、残疾状况和居住在大都市地区的虚拟变量等一系列信息作为控制变量。

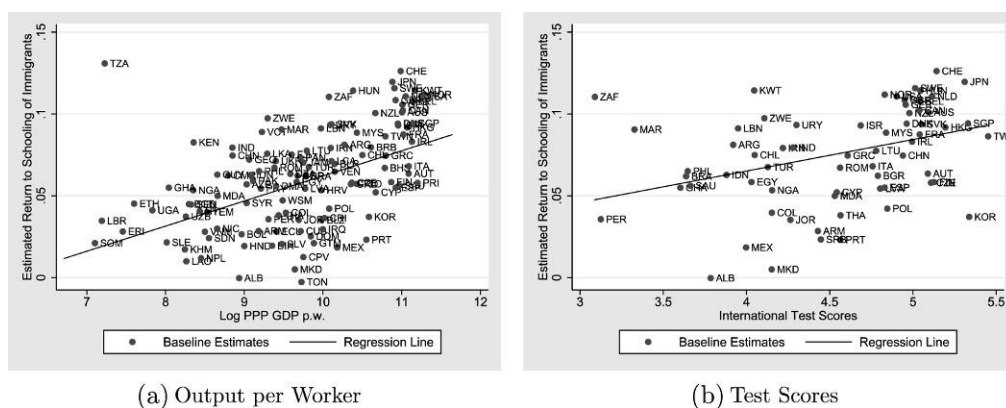


图 A5-1 收益率与人均产出、各国的教育质量

⁴⁸ 实际为统计实体，有些国家在分裂前分裂后都被计算在内。

利用数据估计出来的教育收益率和人均产出呈明显的正相关。并且教育回报率与教育质量的另一个衡量指标——国际标准化成就测试项目的分数相关；使用其他的数据，如 1990 年的美国人口普查，加拿大 2001 年的人口普查，结果在数量上是稳健的（Quantitatively Robust）；因此其得到的收益率能够代表教育质量。

在样本选择性问题上，文章认为其结果并不受选择性的影响⁴⁹，其一是因为文章只关注 γ_{US}^e ，不关注 γ_{US}^e ， γ_{US}^e 可能与移民的选择或其他与教育质量无关的因素的影响。其二使用没有选择性的以难民和被收容者进入美国的移民做分析，估计的教育回报率与基准回归结果在数量上是相似的（Quantitatively Similar）。

另外，文章认为其结果也并不受技能可转移性的影响，其一，具有更严格的技能导向（Skilled Oriented）移民政策的加拿大的估计结果与美国的基准回归结果是相似的；其二，区分移民时间是否达到 15 年、英语表达是否非常流利的分样本回归结果与基准回归结果数量上是相似的（Quantitatively Similar）；其三，其区分工作岗位是否需要技能证书后，两类人群的结果呈现数量上相似（Quantitatively Similar）。

（三）宏观路径下的人力资本价格：从移民工资分离各国的教育质量信息。

Martellini et al. (2024) 利用人力资本价格这一方法估计了 2800 多个大学的毕业生教育质量（大学加总后得到各国的大学教育质量）。文中对于大学毕业生移民而言，工资公式如下：

$$\log(w_{i,c,j,t}^{b(j)}) = \log(z_c) + \log(h_{i,c,j}^{b(j)}) + \eta_{i,c,j,t}^{b(j)}$$

与之前不同的是，公式多了下标 j ，表示大学； $b(j)$ 表示大学所在的国家； t 表示数据时间； $h_{i,c,j}^{b(j)}$ 为个体的人力资本，当在学校层级进行平均后，就可以获得可以衡量学校的质量的指数，此时，公式变为：

$$\log(w_{i,c,j,t}^{b(j)}) = \log(z_c) + \log(q_{i,c,j}^{b(j)}) + \varepsilon_{i,c} + \eta_{i,c,j,t}^{b(j)}$$

当知道技能价格 z_c 后，可以用学校固定效应获得。

与 Schoellman (2012) 等研究不同的是，Martellini et al. (2024) 试图将移民

⁴⁹ 需要注意的是对于文章的稳健性结果应该慎重看待，其稳健性验证的结果虽然与基准回归结果存在较大的相似性，而忽略相似性并不代表完全没有差别，常常一个国家在主要结果中高于另一个国家，在稳健性时却低于这个国家，这问题的存在对于衡量教育质量的努力是不小的损害。

选择和技能可转移性包涵在模型之中。在技能可以完美转移和不存在移民选择的情况下，一个从 $b(j)$ 移民至 c 国家的大学毕业生的收入变化为：

$$\log(w_{i,c,j,t'}^{b(j)}) - \log(w_{i,b(j),j,t}) = \log(z_c) - \log(z_{b(j)}) + \eta_{i,c,j,t'}^{b(j)} - \eta_{i,b(j),j,t}$$

$$\text{其中 } E(\eta_{i,c,j,t'}^{b(j)} - \eta_{i,b(j),j,t} | z_c, z_{b(j)}) = 0$$

考虑技能并不是完美转移，在移民过程中会因技能不适配而产生技能折损：

$$\log(h_{i,c,j}^{b(j)}) = \log(h_{i,b(j),j,t}) + \tau_{b,c}^h, \text{ 此时 } \tau_{b,c}^h < 0;$$

考虑移民选择，一些其他特征会促使某些大学毕业生选择移民，使得在移入国家获得更高的工资：

$$E(\eta_{i,c,j,t}^{b(j)} | z_{b(j)}, z_c) = E(\eta_{i,b(j),j} | z_{b(j)}, z_c) + \tau_{b,c}^\eta, \text{ 此时 } \tau_{b,c}^\eta > 0。$$

由此，大学生移民的收入变化为：

$$\begin{aligned} \log(w_{i,c,j,t'}^{b(j)}) - \log(w_{i,b(j),j,t}) \\ = \log(z_c) - \log(z_{b(j)}) + \tau_{b,c}^h + \tau_{b,c}^\eta + \tilde{\eta}_{i,c,j,t'}^{b(j)} - \tilde{\eta}_{i,b(j),j,t} \end{aligned}$$

$$\text{其中 } \tilde{\eta}_{i,c,j,t'}^{b(j)} \equiv \eta_{i,c,j,t'}^{b(j)} - \tau_{b,c}^\eta, \text{ 所以 } E(\tilde{\eta}_{i,c,j,t'}^{b(j)} - \eta_{i,b(j),j,t} | z_c, z_{b(j)}) = 0$$

这一公式表明，移民收入的变化不仅取决于来源国（Origin）和移入国（Destination）之间的技能价格变化，还来自技能可转移性效应（这一效应为负）和移民选择效应（这一效应为正）。因此，最终 $\tau_{b,c}^h + \tau_{b,c}^\eta$ 的正负号并不确定。

为了对上式估计，[Martellini et al. \(2024\)](#) 利用下式来估计 $\tau_{b,c}^h + \tau_{b,c}^\eta$ 的联合影响：

$$\tau_{b,c}^h + \tau_{b,c}^\eta = \begin{cases} d_c + \beta|y_c - y_b| & \text{if } c \neq b, \\ 0 & \text{if } c = b. \end{cases}$$

简单而言，技能可转移性效应和移民选择效应由目标国家的移民政策和制度（ d_c ）以及两个国家在经济发展水平上的差异（ $\beta|y_c - y_b|$ ）所导致。

由此，最终的估计公式：

$$\begin{aligned} \log(w_{i,c,j,t'}^{b(j)}) - \log(w_{i,b(j),j,t}) \\ = \log(z_c) - \log(z_{b(j)}) + d_c + \beta|y_c - y_b| + \hat{\eta}_{i,c,j,t'}^{b(j)} - \eta_{i,b(j),j,t} \end{aligned}$$

为了获得在技能可转移性效应和移民选择影响下的各国技能价格 z_c 估计，[Martellini et al. \(2024\)](#) 通过利用了特殊的数据。其数据来源于网上求职平台 Glassdoor 网站，网站拥有 2800 所大学毕业生在移民前和移民后的工作信息，并且不同与之前数据只有移民至美国的样本，网站拥有从各个国家移民至各个国家的样本。

在这一数据下，[Martellini et al. \(2024\)](#) 可以直接估计包含考虑了样本选择和技能可转移性的模型。如前所述，估计策略为两步，第一步利用不同大学毕业生在不同国家（大学毕业国家和移入国家）的收入使用公式估计各国国家对收入的影响，获得各国的技能价格 z ；第二步是在获得各国的技能价格情况下，利用学校固定效应，获得各学校的教育质量。

作者最后的结果表明最富裕国家的大学毕业生比最贫穷国家的大学毕业生拥有 50%以上的人力资本。移民加剧了这些差异：来自较贫穷国家的移民往往拥有较高的人力资本水平。