# 一国人力资本究竟应如何测量？

## 从受教育年限到认知技能与市场工资信息

（网络版）

Version: 2024-02-25

云如先

（南京大学教育研究院 210023）

# 说明

此为论文《一国教育人力资本究竟应如何测量？》的网络版。

与期刊版相比，网络版主要在以下几个方面有所不同：一是相关概念，如在网络版中为教育人力资本，网络版中为人力资本。二是网络版中包含了因版面限制、便于理解所缩减的内容。三是总结和评述部分为第一作者原先的粗糙想法。最后但最终要的一点是网络版附录中给出了作者对以往文献方法的详细分析和总结（作者正是在这些资料的基础上编写而成的论文），读原文献较为困难或者想节省精力的读者可以阅读附录。

此外，由于能力、视野所限，若读者发现文章有任何遗漏、错误或不当之处，欢迎指正，作者将谦虚接受并及时更正、更新此文档。也欢迎读者提出自己的想法、对文章进行补充、对文章内容进行更新。我将在特定部分对为本文做出贡献的人做出感谢。

由于是网络版，作者在用词上略微放松，非学术化一些，还请包容。

免责声明：（1）本文为作者的知识分享，用于公益目的。（2）禁止（其他公众号、网站等）转载。（3）本文应当不涉及任何权利问题。若有问题，请与作者本文联系。

请引用：

作者联系邮箱：jssyyrx@163.com。

本文目前可以从以下网址获得：

# 感谢

感谢北京大学教育经济学博士生学术论坛外审专家等提出的宝贵意见。感谢南京大学教育研究院何沛芸、刘铖、廖彬、王玲、吴凯霖对本文修改过程中所作的贡献。

# 一国人力资本究竟应如何测量？

## 从受教育年限到认知技能与市场工资信息

云如先 黄 斌 祝雅汶

（南京大学教育研究院 210023）

摘要：人力资本作为宏观经济研究中的核心概念，拥有人力资本的精确测量是进行宏观经济研究的提前，对衡量人力资本的指标和数据有充分的了解是也是学者进行相关研究的重要基础。为了加深对现有人力资本测量指标和数据的了解，本文沿着人力资本测量指标的“演变史”，介绍了当前常用的人力资本测量指标以及存在问题、数据构建原理和方法，这些人力资本测量指标主要包括受教育年限[1]、学生认知技能、经质量调整后的受教育年限和成人认知技能和从市场工资信息分离出的教育质量和人力资本。依据人力资本理论与教育生产函数理论所提供的统一框架，对各指标的分类和特点进行介绍，提出评判人力资本测量优劣的五个基本原则，即“全面测量优于单维测量”“直接测量优于间接测量”、“存量测量优于流量测量”、“产出测量优于投入测量”和“质量测量优于数量测量”，在这一标准下，成人认知技能应当是最优的测量指标。最后，本文总结人力资本测量指标的总体发展规律，评价指标以及数据存在局限，最后对未来人力资本研究的发展方向进行了展望。

关键词：人力资本；受教育年限；认知技能；市场工资信息

基金资助：江苏省研究生科研与实践创新计划项目“出生队列下教育发展、教育不平等和收入不平等关系研究”（KYCX24_0095）。

[1] 受教育年限、学生（成人）认知技能是个体微观统计量，在另一角度也可以看作是数据的类型，而平均受教育年限、学生（成人）认知技能均值为总体宏观统计量，在另一角度也可以看作是数据的应用。然而在宏观研究中，在提及学生（成人）认知技能时，通向指向的就是学生（成人）认知技能均值，这是语言上的习惯，并且会对理解产生阻碍。为此，本文尽可能在用词上将其区分开。这种区分主要原因在于，一来可以将微观和宏观区分并串联起来；二来均值只是一个总体统计量，其他如教育 GINI 系数、认知技能偏态等总体统计量也可以从微观个体统计量中获得，只是在受教育年限上，平均受教育年限使用最多，其他使用较小。然而本文虽然有意对此进行区分，但受能力所限，读者也应当知晓相应语境下的含义。

# How to measure a Country's human capital?

# From Years of Schooling to Cognitive Skills and Wage Information from Market

Ruxian Yun, Bin Huang, Yawen Zhu

(School of Education, Nanjing University, Nanjing 210023, China)

**Abstract:** As the core concept of macroeconomic research, the accurate measurement of human capital is the advance of macroeconomic research, and having a full understanding of the indicators and data to measure human capital is also an important basis for scholars to conduct related research. In order to deepen the understanding of the existing human capital measurement indicators and data, this paper introduces the current commonly used human capital measurement indicators, as well as the existing problems, and data construction principles and methods , along the "evolution history" of human capital measurement indicators, which mainly include years of education, students' cognitive skills, quality-adjusted years of schooling, adult cognitive skills, and education quality and human capital separated from market wage information. Based on the unified framework provided by the theory of human capital and the theory of educational production function, this paper introduces the classification and characteristics of each measurement, and puts forward five basic principles for judging the advantages and disadvantages of human capital measurement, namely, "comprehensive measurement is better than single- dimension measurement", "direct measurement is better than indirect measurement", "stock measurement is better than flow measurement", "output measurement is better than input measurement" and "quality measurement is better than quantity measurement". Finally, this paper summarizes the overall development law of human capital measurement indicators, the limitations of evaluation indicators and data, and finally looks forward to the future development direction of human capital research.

**Keywords:** Human Capital; Years of Schooling; Cognitive Skills; Wage Information from Market

# 一、引言

人力资本是推动国民经济增长最为重要的因素之一，因此要完成一项宏观经济计量研究或其他以经济社会发展为主旨的研究[2]，必须先实现对一国的人力资本的精确测量。尽管人力资本投资有教育、迁移、健康等多种途径，但教育是实现一国人力资本累积最为重要和主要的途径，因此在实际研究中，人力资本常被狭义地限定为单纯来自教育投资，故而以往研究中常只使用教育相关指标来测量[3]人力资本（Reiter et al., 2020）[4]。准确来说，此种操作所得到的结果更多指向的是教育人力资本，而非完整的人力资本。

二十大报告提出“教育、科技、人才”三位一体发展战略，充分体现了教育对推动强国建设的基础性、战略性作用。教育强国战略强调教育对外部社会经济发展的生产性功能，要求教育要提高对实现共同富裕与中国式现代化战略目标的支撑度与贡献度。一直以来，发展教育被认为是国家获取和累积人力资本最重要的投资方式，而人力资本又是促进长期经济增长的重要源泉(Lucas, 1988；Romer, 1986；Schultz, 1961)。因此，自然而然地便产生了一个最基础和最重要的问题：一国人力资本究竟该如何测量？

人力资本的测量问题是宏观政策与计量研究的基础性工作，如果我们不能解决有关人力资本的测量问题，保证其测量精度和效度，就无从谈起后续研究的效度。已有研究指出，以往宏观研究无法得到令人满意的结论在很大程度上与使用的人力资本存在较大的测量偏误有着密切的联系（Krueger & Lindahl, 2001；De La Fuente & Doménech, 2006）。另一方面，学界对于国别间的人力资本测量指标数据向来实行地是“拿来主义”，对所采用的人力资本测量指标构建详情知之甚少，因而常错误地将各国原本不可比的教育人力资本数据“强行”进行分析，由此得到的国际比较结果往往并不可信。

此外，近年来，尤其是 2010 年后，人力资本测量指标也取得了巨大的发展，除了传统的受教育年限外，出现了学生认知技能、质量调整后的受教育年限、成人认知技能等指标，被应用于一些研究之中，取得了具有重要政策含义的结论。以上情况都使得对已有人力资本测量指标的构建方法进行系统回顾与

[2] 主要为发展核算（Development Accounting）、增长核算（Growth Accounting）以及实证的宏观增长方程（Empirical Macro Growth Equations）研究。

[3] 本文并没特别区分衡量、测量、度量三词的用法。

[4] 衡量教育系统常被认为就是在测量人力资本，这导致了一些测量和相关术语的混乱，如测量教育系统的质量被认为是测量人力资本的质量。

对比审视变得迫切和重要。然而，据作者所知，目前中国还尚无研究聚焦此问题，国外零星文献对此的讨论也不全面（De La Fuente & Doménech, 2024）。

有鉴于此，本文先对国家教育人力资本测量指标的演变历程进行梳理（第二章），继而沿着发展脉络对常用的人力资本测量指标进行系统介绍（第三到七章），包括各种指标的数据存在问题（什么问题？）、数据构建原理和方法（怎么解决？）。最后总结对这些指标的一般性认识和总体的发展规律，并依据人力资本理论与教育生产函数理论构建统一的分析框架，将本文提及的所有指标放置于这一框架下（其八章）。最后总结人力资本测量指标的发展规律，评价其存在的局限，并对未来国家人力资本研究的发展方向进行了讨论（第九章）。希望借助本文，可以加强研究者对对教育人力资本测量指标和数据的了解，帮助读者更好地将这些相关指标与数据运用于研究实战[5]。

## 二、国家人力资本测量指标的“演变史”

自 20 世纪 60 年代，联合国教科文组织系统收集了各国各层级教育的入学率等数据，因而早期宏观经济计量研究常使用入学率来代理各国人力资本，将其用于实证分析人力资本对总量生产率影响的研究之中（Hanushek & Kimko ,2000；Barro & Lee, 2013）[6]。入学率指标的好处在于易于获得，但很明显，它与人力资本理论内涵存在偏差：人力资本是一种存量概念，而入学率是一种教育流量指标，人力资本对入学率的反应是渐进的，具有非常大的滞后性，故而一个国家某一时期的入学率并不能放映该国当期成人劳动力的人力资本存量累积水平。相较而言，从理论的角度看，受教育年限作为教育产出指标，因而一国的平均受教育年限（Mean Years of Schooling，MYS）比入学率这种投入指标更能反映各国的人力资本，并且对一国成人劳动力人口的受教育年限测量，计算该国的平均受教育年限，其实就是对该国人力资本存量进行测量。在这一理念的驱动下，国际组织和大量研究者开始尝试构建起各国劳动力平均受教育年限数据库，以实现对各国人力资本存量更为直接的测量，平均受教育年限也逐渐替代入学率，成为衡量一国人力资本的常用指标（De La Fuente & Doménech, 2006）。

平均受教育年限作为一国人力资本测量指标也存在不少缺陷，随着这一指标的广泛使用，先后出现了两个批评。第一个批评强调不同来源提供的平均受

[5] 本文也给了各类方法的文献汇总以及常见的数据库结果，详见附录。

[6] 在早期，识字率（literacy rate）也被大量地应用于实证研究之中，但相比较而言，其更难获得。有关识字率和入学率的讨论详见 Berro & Lee（1993）。

教育年限数据都存在较大的测量误差，这使得以平均受教育年限作为一国人力资本存量测量指标对经济增长的实证分析经常得到令人沮丧的结果（Krueger & Lindahl, 2001；De La Fuente & Doménech, 2006）。第二个批评指出平均受教育年限只测量了一国国民接受教育的数量，而忽略了各国教育质量的巨大差异。宏观计量研究常使用国家和地区作为分析的样本，而不同国家和地区的教育质量显然存在较大的差异，这意味着在不同国家接受相同年限的教育所产生的人力资本理应是不同的，而以受教育年限来测量人力资本，隐性地假设在不同国家和地区接受相同年限的教育能获得相同的教育结果，这明显与事实不符。比如，事实上很少有人会认为美国中学的一年教育和埃及的是等价的（Hanushek & Kimko, 2000）。这也就是说，即使两个国家的平均受教育年限是一样的，但由于两个国人口所接受的教育质量存在巨大差距，那么这两个国家的人力资本存量也就存在巨大差异（事实上，无人敢说两个国家任何时期的教育质量都是一样的）。基于此观点，平均受教育年限作为一国教育数量指标，显然不能很好地反映一国的人力资本。

尽管第一种批评随着数据的改善逐渐减少，第二种却是日益强烈。研究者们意识到，理想的人力资本测量指标应既要反映一国人力资本的数量，也要很好地测量出一国人力资本的质量。为在研究中控制各国教育质量差异，不少研究使用师生比、生均经费等指标来代理各国的教育质量，但这些指标都是教育生产投入要素，将这些投入都看作是有效的生产性投资是不合适的（Hanushek, 2003），投入能够产生多大的产出也是存在不确定性的（Schoellman, 2012; Hanushek, 2003）。有关“教育投入有用还是无用”，时至今日学界仍存在巨大争议。例如，与城市学校相比，农村学校的生师比都偏小，但农村学校的教育质量通常要比城市学校差；同理，受规模效应的影响，小规模学校的生均成本（经费）都比较高，但小规模学校的教育质量未见得要比其他学校高。

相对而言，从教育结果来测量一国教育质量与人力资本具有学理上的巨大优势。伴随着“去学校不等同于有学习收获（Schooling is not Learning）”的提倡（Pritchett, 2013），“（学生）认知技能”脱颖而出，进入研究者们的视野。人力资本理论认为一个人接受教育是为了获取技能，而这能使得个人劳动生产率提升，进而有助于个人收入提高和国民经济增长。因而，认知技能被引入对人力资本的测量之中，这被认为是人力资本理论概念的回归（Hanushek & Woessmann, 2015；黄斌 等，2024）。学生认知技能首先被众多研究所使用，这是因为国际性和地区性组织在不同国家广泛地开展学生能力评估测试项目，这些测试项目测量了数学（Math）、文学（Literature）、科学（Science）等多方面

的素养，其各素养得分可以很好地衡量学生的认知技能。鉴于不同测试项目采用的试题有较大的差异，为获得更大的数据样本，需要实现不同学生认知技能测试得分之间的可比化转换，不少学者致力于解决这一问题，涌现出一些非常有价值的方法（Hanushek & Woessmann, 2012; Angrist et al., 2021; Gust et al., 2024）。

尽管学生认知技能作为教育质量衡量指标日益受到重视，并且相关研究发现同时控制学生认知技能均值和平均受教育年限时，只有学生认知技能均值显著影响一国的经济增长[7]，但平均受教育年限作为教育数量的衡量标准，并未因此而被完全淘汰。这一方面是由于受教育年限在广泛的被接受程度上具有无可比拟的优势；另一方面是因为学界从未否定过教育数量的重要性，认为教育数量和质量同等重要。在这种情况下，涌现出一系列研究，其意图通过一定方法使得教育数量和教育质量相结合，构建既包含教育数量信息，又包含教育质量信息的新指标。这类研究通常从学生认知技能数据中估计教育质量，参照某一基准（国家），计算质量调整系数，对平均受教育年限进行调整，获得质量调整后的受教育年限（Filmer et al., 2020）。

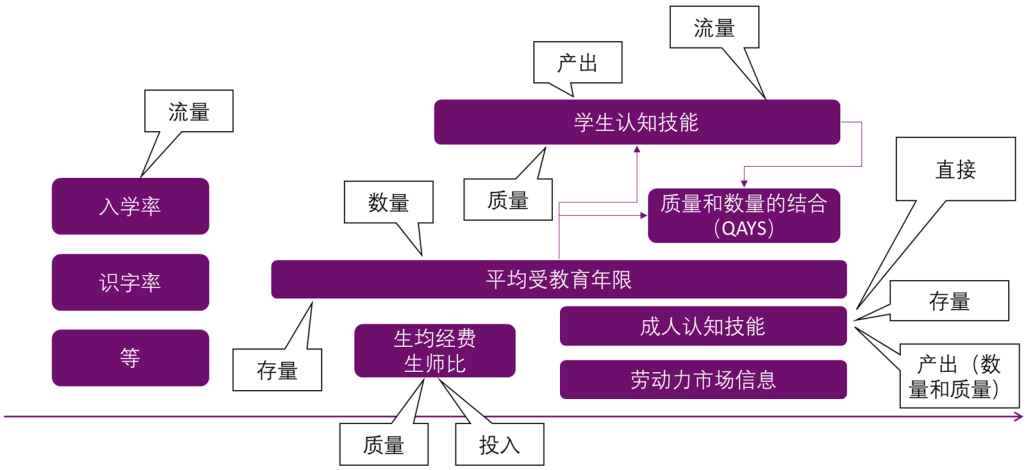
同时，过多地使用学生认知技能又陷入了和入学率一样的以流量代替存量的缺陷之中；并且理论上，当前国家的人力资本存量应当受人口所对应学生时代的教育质量的影响，并不会受当前国家的教育质量的影响[8]。由于使用学生认知技能所产生的一系列问题都可以通过直接使用成人认知技能数据所解决。此外，成人认知技能数据兼具教育产出与存量两种优良品质，并且根据人力资本概念，测量认知技能可以当作是对人力资本的直接测量，因此成人认知技能自然而然地成为了人力资本测量指标的最优选择。然而成人认知技能测试项目的稀少和包含国家的有限性，为突破这一局限，有研究根据流量和存量之间的关系，试图通过建立学生认知技能和成人认知技能的关联，来构建包含更多国家的成人认知技能数据，这一方法同样具有开创性（Égert et al., 2024）。虽然目前该方法尚处于起步阶段，但未来有着巨大的应用价值，非常值得关注。

此外，也有一些学者在构造新的人力资本测量指标上采用不同的处理方法。如，一些研究另辟新径，完全逃脱教育投入-产出的分析框架，基于劳动力

[7] 有学者认为这是因为平均受教育年限和学生认知技能均值之间巨大的共线性所导致，而非教育数量无用（Hanushek & Woessmann, 2012）。

[8] 除理论上的不合适外，已有经验证据表明，学生与成人认知技能的各特征虽然具有一定的相关性，但尤其是分布偏态、标准差等特征上存在着较大差异（黄斌 等，2024）。这些都促使学者反思使用学生认知技能数据替代成人认知技能数据的研究。

市场的工资信息是对人力资本的直观反映，从工资信息中分离出各国的教育质量和人力资本，提出了颇具创新性的方法（Schoellman, 2012; Martellini et al., 2024）。该方法目前虽然受制于市场上工资信息获取问题等原因，在教育质量和人力资本的变化、衡量增量方面存在不足，但其价值却不容忽视。




注：时间随着坐标轴从左到右逐渐推进。

**图 2-1 人力资本测量指标发展大致脉络**

接下来，本文将沿着国家人力资本测量指标发展脉络，对它们的存在问题、数据构建原理和方法、现有公开数据等内容进行系统介绍，为之后统一框架的建议、对比分析各类教育人力资本测量指标优劣、总结归纳评判教育人力资本优劣的基本原则奠定基础。

## 三、受教育年限：一种有关教育获得[9]数量的测量指标

截止目前，受教育年限依然是当前最盛行的人力资本指标，同时平均受教育年限也依然是宏观研究中一国人力资本最常用的衡量指标。学界普遍认为平均受教育年限能够衡量一国教育获得的数量。

### （一）补充缺失年份的数据：受教育年限的插补问题

一般来说，各国的平均受教育年限是由各国的人口普查或调查的个体微观数据计算而得。该指标在计算上存在两个问题：

一是各国教育层级与类型分类的不统一。一方面，各国教育系统和学制制

[9] 需要提醒的是，在英文中，教育获得（Educational Attainment）通常指教育水平或受教育年限，其内涵本身就偏向数量维度，并没有过多涉及质量。

度不同，各个国家有不同于其他国家的教育系统，包涵不同类型的学校，并且在不同国家完成看似相同的教育水平所需的年限可能存在一定差异；另一方面，即使是同一国家，不同时期其学制也可能会发生变化。要解决这个问题，就需要对样本中每一个国家的教育系统与学制制度做细致地对比分析。此项工作并不涉及复杂技术，但由于繁琐复杂，需要耗费大量时间和精力进行整理。

二是各国的人口普查通常是每 10 年进行一次，时间间隔太长，若要形成间隔更短（如 5 年一次）的受教育年限面板数据，就需要对无人口普查或调查的间隔年份进行数据插补。为实现数据插补，以往学者提出了若干方法。早期研究提出的方法相对简单，包括永续盘存法（Perpetual Inventory Method）（Barro & Lee ,1993; Barro & Lee, 2001）、简单线性插补（Simple Linear Extrapolation）（De La Fuente & Doménech, 2000； De La Fuente & Doménech, 2006）等[10]。2007 年，研究者开始使用人口出生队列信息进行数据插补，提出出生队列趋势外推法（Forward and Backward Extrapolation）[11]（Cohen & Soto, 2007; Barro & Lee, 2013; De La Fuente & Doménech, 2015; Barro & Lee, 2015）和出生队列迭代后推法（Iterative Multi-dimensional Cohort-component Reconstruction）（Lutz et al., 2007; Bauer et al., 2012; Goujon et al., 2016; Speringer et al., 2019）。

当今国际宏观研究中最常用的数据库是基于 Barro & Lee (2013) 的方法而构建的，使用的是出生队列趋势外推法。为了加强对平均受教育年限的认识，综合考虑方法的发展历史和应用性，本文将将对主流（曾）使用的永续盘存法、出生队列趋势外推法与出生队列迭代后推法都进行介绍。这三种方法原理各有不同，但互有联系：永续盘存法是利用基期数据对之后年份进行数据插补，迭代后推法是利用最新的数据作为基期对之前的年份数据进行数据插补，而趋势外推法则是基于基期数据同时进行前推与后推。

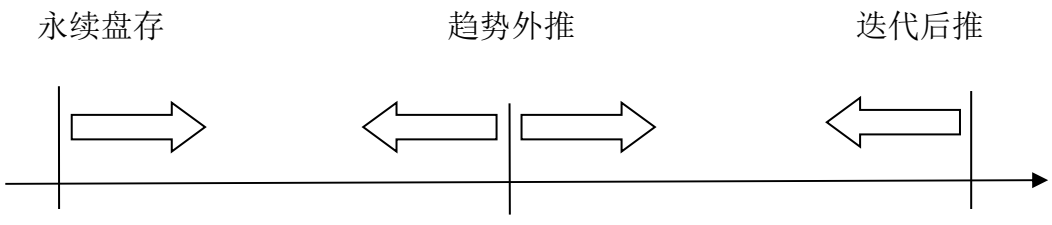



**图 3-1 三种数据插补方法示意图**

[10] 相较而言，简单线性插补原理较为简单，本文对此方法就不做介绍。关于早期的受教育年限插补方法，可参见 De La Fuente & Doménech（2006）。

[11] 这系列文章并没有给其方法命名，我们摘取了关键词作为其方法的名称。

## （二）受教育年限的插补方法

### 1. 永续盘存法

永续盘存法是以某个国家某一时期的人口普查或调查数据作为基期，通过考虑由于死亡和新增入学人口所导致的后续时期各教育阶段人口[12]的变化，以此实现对该国后续年份的受教育年限数据补值。

通常情况下，各国人口的平均受教育年限由各教育阶段人口占比与获得该教育阶段所需的时间相乘并加总而得[13]，即有：

$$ys_t = \sum_{l=0}^{l} h_{l,t} Dur_{l,t} \quad （3\text{-}1）$$

$$h_{l,t} = \frac{H_{l,t}}{L_t} \quad （3\text{-}2）$$

其中，$ys_t$为第$t$年某一国家人口的平均受教育年限；$l$表示各教育阶段，通常使用较为宽泛的教育层级划分：未受教育时$l = noedu$，完成初等教育时$l = pri$，完成中等教育时$l = sec$，完成高等教育时$l = ter$；$h_{l,t}$表各教育阶段的人口占比，它等于各教育阶段$l$的人口数（$H_{l,t}$）除以总人口（$L_t$）；$Dur_{l,t}$为在该国为达到该教育阶段所需的教育年限。

假设我们拥有基期的各教育阶段人口数（$H_{l,t-5}$），要对五年后的各教育阶段人口数（$H_{l,t}$）进行补值。根据永续盘存法，可采用如下公式：

$$H_{l,t} = H_{l,t-5}(1 - \delta_t) + add_{l,t-5\ to\ t} \quad （3\text{-}3）$$

其中$\delta_t$为死亡率，$H_{l,t-5}(1 - \delta_t)$为经死亡调整后的各教育阶段存活的人口，

---

[12] 为了统计的完整性，有两种统计方式。第一种是通常使用的，也是本文及引用文献所使用的，其对按最高学历对所有人口进行竖向的分割，此时各较教育阶段人口为最高学历达到该教育阶段的人口（如，未受教育的人口、小学学历的人口、初中学历的人口等等），其对应的受教育年限为获得对应教育阶段所需的所有时间，比如小学对应 6 年、初中对应 9 年、高中对应 12 年等等；第二种是对人口进行横向的分层，此时各较教育阶段人口为完成该教育阶段的人口（如未受教育人口、上了小学的人口、上了小学后又上了初中的人口、上了初中又上了大学的人口等等），其对应的受教育年限为获得完成该教育阶段所需要的时间，比如小学对应 6 年、初中对应 3 年、高中对应 3 年等等。事实上，在获得各国受教育人口结构之后，两种统计方式都能得到一致的结果。

[13] 现有平均受教育年限数据构建通常为 15 岁-64 岁人口，为了简单，后文不再对年龄进行强调。

$add_{l,t-5\ to\ t}$为新增的各教育阶段人口的变化，通常使用入学率信息进行测算。

该公式可以直观地理解为各教育阶段人口等于基期完成该教育阶段的人口减去该教育阶段的死亡人口，加上完成该阶段教育的新增人口。举例而言，某一国家在 2010 年进行人口普查，于是根据以上公式（3-1）-（3-2）可计算出2010 年人口的平均受教育年限，但如果我们需要该国 2015 年人口受教育年限数据，就可以应用公式（3-3），即计算下式：

$$H_{l,2015} = H_{l,2010}(1 - \delta_{2010}) + add_{l,2010\ to\ 2015}{}^{14} \quad （3\text{-}4）$$

由上式可知，死亡率是不随出生队列（年龄）和教育阶段而变化的，这被认为是永续盘存法的最大问题，因为现实往往是教育程度越高，死亡率越低（Balaj et al., 2024）；年龄越长，死亡率越高[15]。有学者怀疑由于这一缺陷可能会产生较大的构建误差，这很可能是使得使相关实证研究无法得到令人满意结果的主要原因（Krueger & Lindahl, 2001；De La Fuente & Doménech, 2006）。

**2. 出生队列趋势外推法**

利用出生队列的信息可以有效降低由死亡率假设导致的测量误差问题（Cohen & Soto, 2007）。在出生队列视角下，人口平均受教育年限的公式如下：

$$ys_t = \sum_{a=1}^{11} l_t^a ys_t^a \quad （3\text{-}4）$$

$$ys_t^a = \sum_{l=0}^{l} h_{l,t}^a Dur_{l,t} \quad （3\text{-}5）$$

式中，$ys_t$为最终总的平均受教育年限，$ys_t^a$为各出生队列的平均受教育年限，$l_t^a$为各出生队列的人口占比，$h_{l,t}^a$为各出生队列中各教育阶段的人口占比。这里引入的新符号$a$为出生队列，各研究采用的出生队列有细微的区别，但通常$a = 1$时表示 15-19 岁，$a = 2$时表示 20-24 岁，以此类推，直到$a = 11$表示 65 岁及以上。与之前不同的是，这里将整个人口按年龄展开为各出生队列，在获得各出生队列的平均受教育年限（$ys_t^a$）后，根据各出生队列的人口占比（$l_t^a$）进行加权加总获得人口的平均受教育年限（$ys_t$）。

[14] 在该方法的应用中，最重要的是估算出各时期的死亡率$\delta_t$，以及使用入学率信息测算新增的各教育阶段人口的变化$add_{l,t-5\ to\ t}$，由于篇幅的限制，本文将不会对此进行详细介绍，如有需要，可参见 Barro & Lee（2001），也可与本文作者联系，以获得由本文作者总结的这些文献方法的具体细节和说明。

[15] 除此以外，移民的发生也会使得不同时间、相同出生队列人群的各教育阶段人口占比发生变化，然后现有研究对此只是略微讨论，尚无研究在方法中对这一点加以考虑。

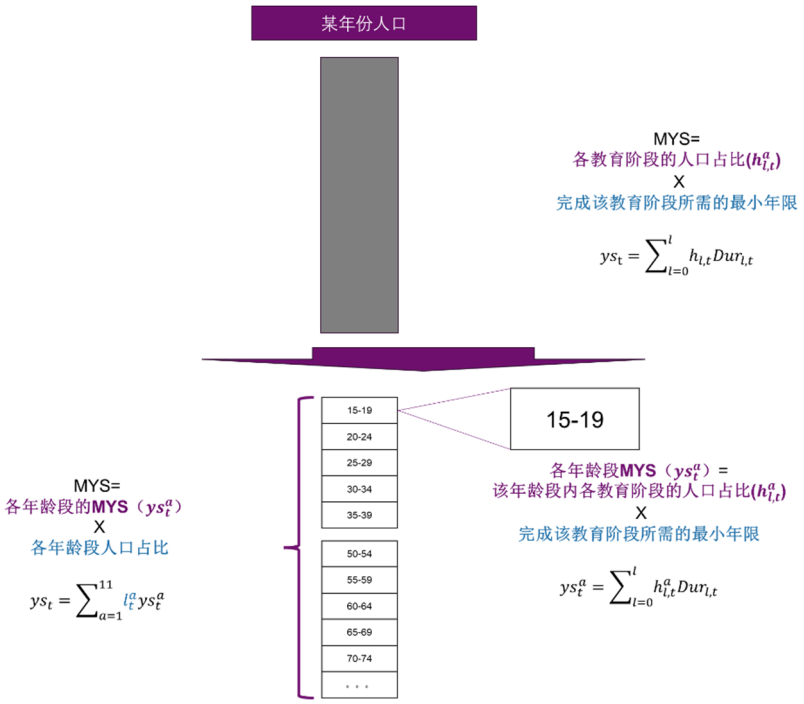



**图 3-2 平均受教育年限计算示意图**

上述公式中，人口信息可以从联合国的经济与社会事务部门人口司的世界人口展望（World Population Prospects，WPP）中获得，可以看作是已知的。因此，若想获得各国平均受教育年限，只需再获得各国各出生队列的各教育阶段人口占比（$h_{l,t}^a$）（Barro & Lee, 2013）或者平均受教育年限（$ys_t^a$）（Cohen & Soto, 2007）中的一个。这两个信息可以使用出生队列外推法获得。

出生队列外推法认为对于一个完成正式教育的人来说，其受教育程度或年限终其一生都不会发生变化，这意味着在同一个出生队列中，各教育阶段人口的占比会始终保持不变，除非该人口出生队列由于死亡发生重大的结构变化（Barro & Lee, 2013）。由于个体的受教育年限终其一生不变，并且同一个出生队列的各教育阶段人口的占比也始终不变，那么这一出生队列人口的平均受教育年限也会保持不变（Cohen & Soto, 2007）。借用这一推论，我们就可以基于基期数据对之前年份和之后年份的人口受教育年限数据进行前推和后推。

需要注意的是，出生队列外推法对于人口死亡率也做了重要假设，它假设同一出生队列人口的死亡率不会随教育程度发生变化，即在同一出生队列，无论受教育程度如何，生存率都是相同的。如此假定，才能保证同一出生队列人口的各教育阶段人口占比不随时间变化（Barro & Lee, 2013），继而才能得到该出生队列人口的平均受教育年限也不随时间变化（Cohen & Soto, 2007）的结论。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，对高龄群体（65 岁以上人口）并不成立，需要对高龄群体进行死亡率调整。另外，对于 25 岁以下低龄人口，因为这些人的受教育情况还在不断发展变化，也需要利用其他方式进行估算[16]。

为更加直观地理解这一方法，我们依据 Barro & Lee (2013)，绘制了图 3。图中，t 期为拥有数据的基期，t+5 和 t-5 为需要插补的年份。在此图中，该方法需要完成下面两个部分内容：

第一，通过拥有数据$t$期的$h_{l,t}^{a}$和$ys_{t}^{a}$，前推后推获得缺失数据$t\pm5$期的$h_{l,t\pm5}^{a\pm1}$或$ys_{t\pm5}^{a\pm1}$。如图中实线箭头为例，通过$t$期的出生队列（30-34）的$h_{l,t}^{4}$和$ys_{t}^{4}$为例，可以通过后推和前推分别获得$t-5$期的出生队列（25-29）的$h_{l,t-5}^{3}$与$ys_{t-5}^{3}$和$t+5$期的出生队列（35-39）的$h_{l,t+5}^{5}$与$ys_{t+5}^{5}$。一般而言，在前推中，可以利用以下两式获得缺失年份的$h_{l,t+5}^{a+1}$或$ys_{t+5}^{a+1}$：

$$h_{l,t+5}^{a+1}=h_{l,t}^{a}\ a=2,\dots,10 \quad (3\text{-}6)$$

$$ys_{t+5}^{a+1}=ys_{t}^{a}\ \ a=2,\dots,10 \quad (3\text{-}7)$$

同样，在后推中，可以利用以下两式获得缺失年份的$h_{l,t-5}^{a-1}$或$ys_{t-5}^{a-1}$：

$$h_{l,t-5}^{a-1}=h_{l,t}^{a}\ a=3,\dots,11 \quad (3\text{-}8)$$

$$ys_{t-5}^{a-1}=ys_{t}^{a}\ a=3,\dots,11 \quad (3\text{-}9)$$

以上公式的作用都是将不同年份间相同出生队列的信息等同。

第二，需要对无法通过趋势外推获得数据的出生队列进行其他方式补值。如图中灰色底纹出生队列（25 岁之前和 64 岁之后的出生队列），在补值处理中通常会考虑入学率、死亡率、移民等因素的综合影响。[17]。

[16] 虽然对此有相同的认识，但 Cohen & Soto (2007)和 Barro & Lee (2013)的公式在出生队列 a 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。

[17] 在几乎所有的文献中，都对入学率和死亡率进行了考虑，但对于移民，只有对其影响的讨论，没有对其的考虑。

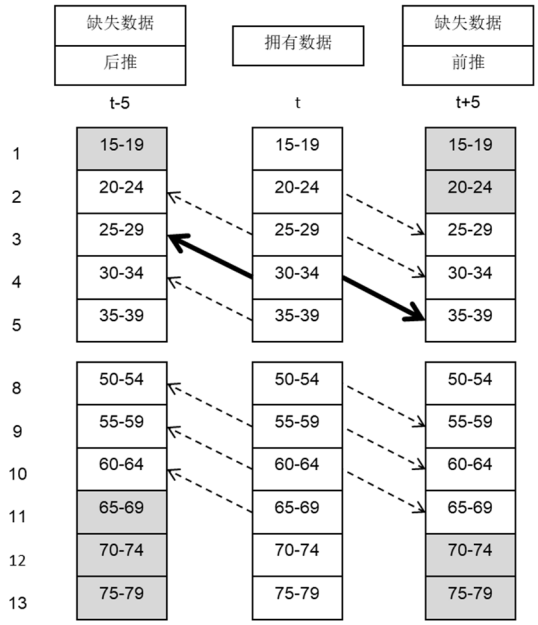



**图 3-3 趋势外推法示意图**

该类研究的一般步骤通常为：首先需要通过趋势外推法（即公式（3-6）-（3-8）或公式（3-7）-公式（3-9））获得可获得信息的出生队列的信息；其次使用一定方法估计无法通过趋势外推法获得信息的出生队列的信息；最后使用公式（3-4）-（3-5）计算缺失数据年份的平均受教育年限。

### 3. 出生队列迭代后推法[18]

与出生队列趋势外推法不同，出生队列迭代后推法[19]只进行后推，不做前推。方法通常选择最新的某一年数据作为基年，基于该年份数据不停地进行迭代后推与补值。实施该方法大致可分为以下两个部分：

第一个部分为数据收集和处理。这部分收集的数据包括：

（1）基年（t）的教育信息数据，这通常来自人口普查。在进行统一的数据预处理后，获得基年 t 的各性别-出生队列-教育阶段的人口占比$h(a,l,t,sex)$[20]。

---

[18] 这一方法所构造的数据在应用上远没有其他方法构造的数据普遍，因此，本文只对这一方法进行简单的介绍。若想进一步了解，见 Lutz et al. (2007) 和 Speringer et al. (2019) 。

[19] 原文将这一方法称作迭代多维队列分量重建(Iterative Multi-dimensional Cohort-component Reconstruction)

[20] 在这一步骤中 Lutz et al. (2007) 是利用的各性别-出生队列-教育阶段人口数量数据，通过各性别-出生队列-教育阶段人口占比和 UN 的人口结构（出生队列-人口数）数据获得人口数量数据。因此在第（4）步的公式对应为人口数量，并在第（5）步后将数量转成占比在结合 UN 的人口结构数据获得新的各性别-出生队列-教育阶段人口数量数据，这一人口数量-人口占比-人口数量转换被人为是对移民情况的考虑。而在 Speringer et al. (2019) 中，则从头到尾使用的都是各性别-出生队列-教育阶段人口占比数据，只在最后人口占比数据结合 UN 的人口结构数据获得人口数量数据。

（2）历年的人口结构数据，与上面几篇文章一致，这一数据来自 WPP。

（3）历年的生命表（Life Table）数据，这一数据同样来自 WPP。该数据主要与不同性别-教育阶段人口的预期寿命差值信息结合，用于计算历年（如$t-5$）的“不同性别-出生队列-教育阶段”人口的幸存率Survival Ratios$(a-1,l,t-5,sex)$，用于幸存率调整。

第二个步骤为迭代计算，为了介绍迭代过程，我们以构建 15-19 至 105+出生队列为例[21]，在图 2-3 中给出了一次完整迭代过程的示意图，图中的文字对应着迭代的步骤和名称。

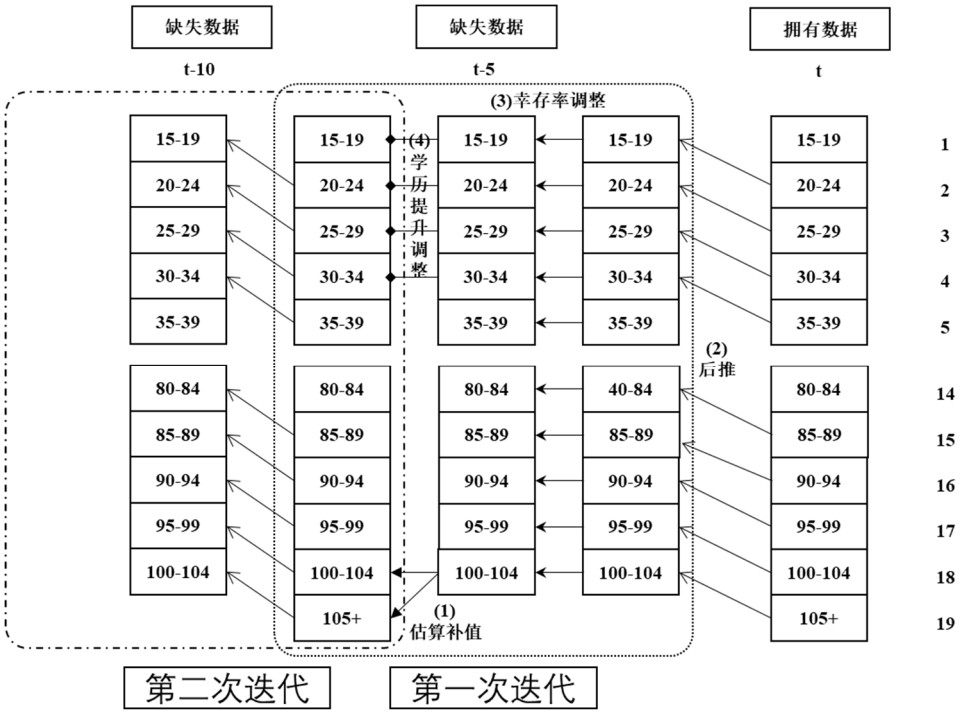



**图 3-4 一次完整迭代示意图**

在图中，迭代计算共有以下几个步骤：

（1）对最高出生队列进行估算补值。如图中所示，在第一次迭代过程中，由于$t$期的出生队列 105+被用来后推$t-5$期出生队列 100-104，这时候对于$t-5$期而言，其最高出生队列（105+）是空缺的，如果要进行第二次迭代，需要先对这一信息进行补值。由于只有在迭代后才会产生最高年龄组数据空缺，才需要补值，因此这一步在第二次及以后的迭代过程中才需要考虑。

[21]在 Lutz et al. (2007) 为 70+；在 Goujon et al., (2016)为 100+；在 Speringer et al. (2019) 为 105+。

（2）后推。这一步和趋势外推法类似，只需要利用以下公式获得 t-5 期的各出生队列的“不同性别-不同出生队列-不同教育阶段”的人口占比数据：

$$h(a-1,l,t-5,sex)=h(a,l,t,sex) \quad (2\text{-}10)$$

该公式同样是将不同年份相同出生队列的信息进行等同。

（3）进行幸存率调整。从$t$到$t-5$过程中，死亡在改变人口结构，因此需要使用幸存率信息Survival ratios$(a-1,l,t-5,sex)$对后推后的结果进行幸存率调整：

$$h'(a-1,l,t-5,sex)=\frac{h(a-1,l,t-5,sex)}{\text{Survival ratios}(a-1,l,t-5,sex)} \quad (2\text{-}11)$$

（4）进行学历提升调整。由于 15-34 岁人口仍存在学历提升的可能性，为此，要对后推得到的 15-34 岁群体进行受教育年限调整。获得最终的$t-5$期的结果。

（5）再回到第 1 步，基于 t-5 期结果，进行第二次迭代计算，以此类推，可以获得$t-10,\ t-15,\ldots$, 期的结果。

上述的（2）和（3）可以合成一步。在该方法的应用中，各文献基本都遵从以上步骤进行迭代补值，不过每部分的阶段性处理略微不同，尤其是在第（2）步、（3）和（5）步的处理上有所差别。

## 四、学生认知技能：一种关注教育获得质量[22]的测量指标

学生认知技能是最近二十年间才出现的指标，学生认知技能均值也逐渐被同于宏观研究之中，并且学生认知技能分布的其他统计量虽然不是重点，但也不少文献涉及。学生认知技能在宏观研究之中常被用作衡量教育获得质量，事实上，对其本身是否能够衡量教育质量，少有讨论，我们将在第八章对此进行讨论。

### （一）构建国际可比数据库：各测试项目得分的可比性问题

近二十年来，随着学界对“Schooling is not Learning”的认识和强调，教育质量的重要性日益凸显。因此，有越来越多的国际性或地区性组织针对各国基

[22] 质量一词在不同的情景中，有不同的含义。比如，一个教育系统有 5 年受教育，其每年教育都可以提升其数学分数 20 分，那么最终这个学生的质量是 100 分，每年的教育质量是 20 分，前者为总体质量、最终质量，后者是单位质量。虽然学生认知技能是一个总产出，但由于在学生年龄或年级一定时，学生认知技能是单位质量。

础教育学生素养开展评估，以此评估各国当前的教育质量，如国际学生评估项目（Programme for International Student Assessment，PISA）、国际数学与科学趋势研究（Trends in International Mathematics and Science Study，TIMSS）、国际阅读素养进展研究项目（Progress in International Reading Literacy Study，PIRLS）、南非和东非教育质量监测联盟项目（Southern and Eastern Africa Consortium for Monitoring Educational Quality，SACMEQ）、法语国家教育系统分析项目（Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN，PASEC）、由拉丁美洲教育质量评估实验室（Latin American Laboratory for Assessment of the Quality of Education，LLECE）实施的第一次、第二次和第三次区域比较研究项目(First/Second/Third Regional Comparative and Explanatory Survey，PERCE/SERCE/TERCE)、早期儿童阅读评估（Early Grade Reading Assessment，EGRA）、年度教育状况报告（Annual Status of Education Report，ASER）、美国国家教育进展评估(United States' National Assessment of Educational Progress, NAEP)、印度国家成就评估（India National Achievement Survey, NAS）等[23]。这些项目测试了能够影响学生未来生产行为的各种认知技能，如数学素养、文学素养（阅读和写作）、科学素养等，这些素养和能力的得分能够很好地衡量各国的学生学习收获，以此可以更进一步地评价各国当时的教育系统质量。

学生认知技能得分在一些宏观研究中被认为是教育质量的最佳衡量指标[24]。然而研究通常想拥有更多国家（横向）和更长时间（纵向）的数据，而各种测试项目在设计上的互不可比，为此实现这一目的需要采用一定方法使得各测试项目在横向和纵向上可比。

具体而言，在横向上，研究想使得相近年份的两个测试项目可比[25]。其目的有两个，一是想包含更多的国家，尤其是包含各个发展阶段和各个地区的国家。然而单个测试项目所包含的国家都相当有限，即使覆盖国家最多的 PISA，也只调查了 102 个国家和地区，这一数目尚不到全球国家数目的一半；并且在经济发展水平上，这些国家更多地是中等偏上收入以上的经济体，中等偏下收入及以下的经济体覆盖较少；在地区上，非洲地区国家数量较少。解决这一问题的办法主要在于使得各个测试项目的得分可比，以此通过整合国际性和地区

[23] 有关这些国际性和地区性测试的简要介绍，见 De La Fuente & Doménech (2024)。

[24] 需要注意的是，这些研究通常没有注意或者忽视时间问题，用现在的教育质量对当前的经济增长做回归，事实上，应当是用当前劳动力市场上人口所对应学生时代的教育质量，这两者有巨大的时间差距。

[25] 多个不同测试项目的转换也是两个、两个的进行，为了表述精确，文中都是使用两个。

性测试项目来包含更多的国家样本，这是现有研究的最主要目的。二是整合两个测试之间的不同信息，实现某一信息的跨地区比较。由于目前这一目的只出现在一类研究中，因此我们将在之后进行详情介绍（见 4.2.2）。

在纵向上，宏观研究也通常想包含更多年份（横向上）的数据，这意味着需要通过一定的方法使得多轮测试之间的成绩可比。在 20 世纪 90 年代之后，各测试项目基本采用项目反应理论，通过相同题项（见 4.2.3）使得测试自身实现纵向时间上的可比，因此并不需要额外的处理方法。而对在 20 世纪 90 年代之前的测试，其相当于将同一测试项目不同轮次的测试当成两个完全不同的测试项目结果，其处理方法与横向国家的处理方法一致，只是在这基础上再加以考虑时间可能导致的整体水平（均值）差异（从这一角度而言，横向和纵向可比是一致的）。考虑到方法相似性以及 20 世纪 90 年代之前数据的特殊性和使用的稀少性，因此，本文中并不会对时间上的可比作过多的介绍。

### （二）锚点与构造转换函数

不论是通过整合国际性和地区性测试项目来扩大国家样本，还是通过整合两个测试的不同信息，都需要先通过锚点来构造转换函数（Transforming or Linking Function），再应用转换函数实现不同测试项目数据的转换。

假设想将测试项目 X 的得分转换成测试项目 Y 的得分，一般化的公式为：

$$Score_Y = f(Score_X) \quad （4\text{-}1）$$

其中，$Score_Y$是我们想得到的测试项目$Y$的得分，$Score_X$是我们已掌握的测试项目$X$的得分，$f(\cdot)$为函数，如果采用简单线性的话，公式可变为：

$$Score_Y = \alpha + \beta * Score_X \quad （4\text{-}2）$$

其中，$\alpha$和$\beta$为转换参数。

利用转换函数，我们既可以实现微观层面学生个体分数的转换，也可以实现宏观层面国家均分或总分的转换，这取决于所使用的方法和数据层次。

构建转换函数的关键在于拥有锚点（Anchoring），锚点是指两个不同测试项目之间的重叠之处。锚点可以是两个测试之间重叠的学生、国家和题目（或题项）。只有拥有锚点才能构建转换公式；可以说，相同个体（Common Persons）、相同总体（Common Populations）或相同题项（Common Items or Overlapping Items）是构建转换函数的基础（Kolen & Brennan, 2014；Reardon et

al., 2021）[26]。当我们运用不同类型锚点时，所采用的技术会有所差别，下图给出了方法的整体框架，我们将对其进行一一介绍。

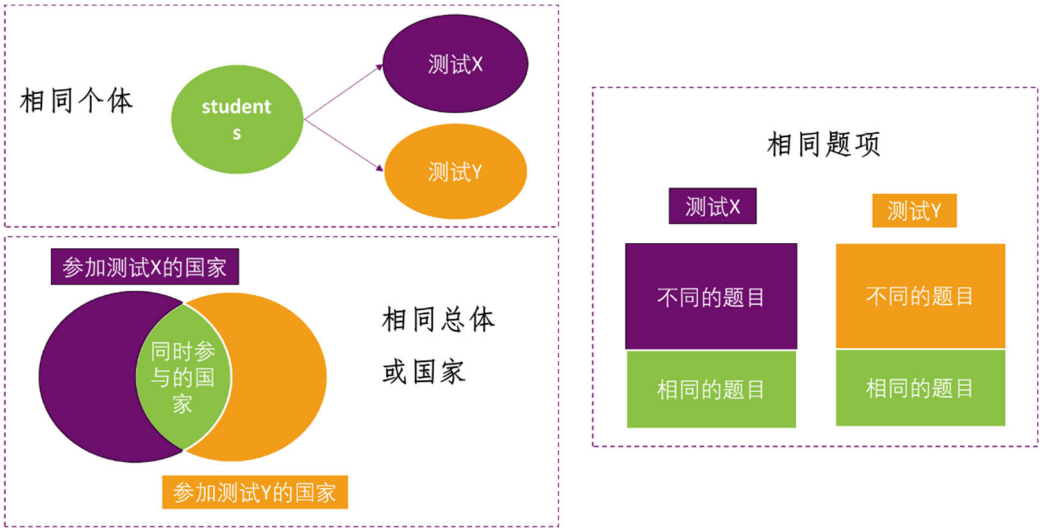
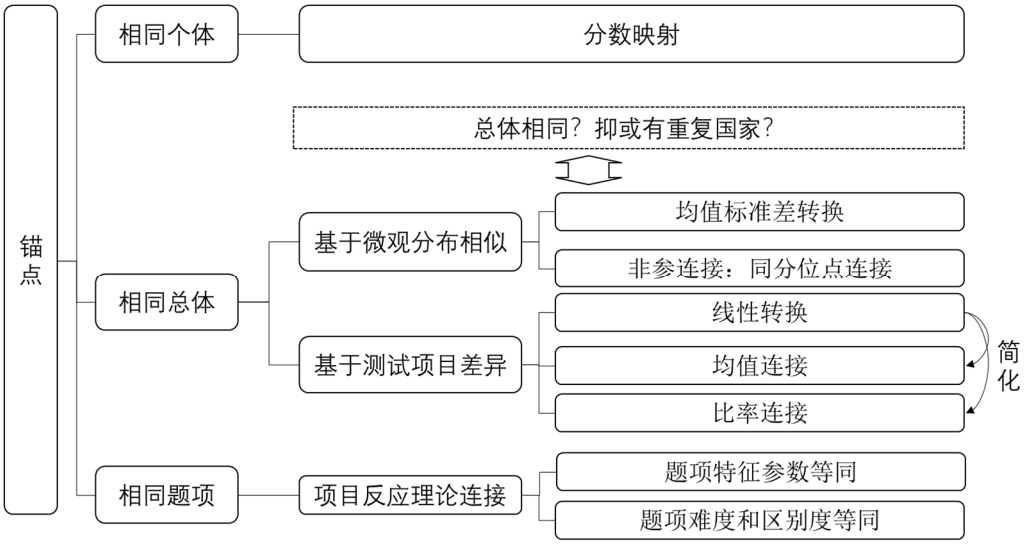



**图 4-1 锚点情况示意图**



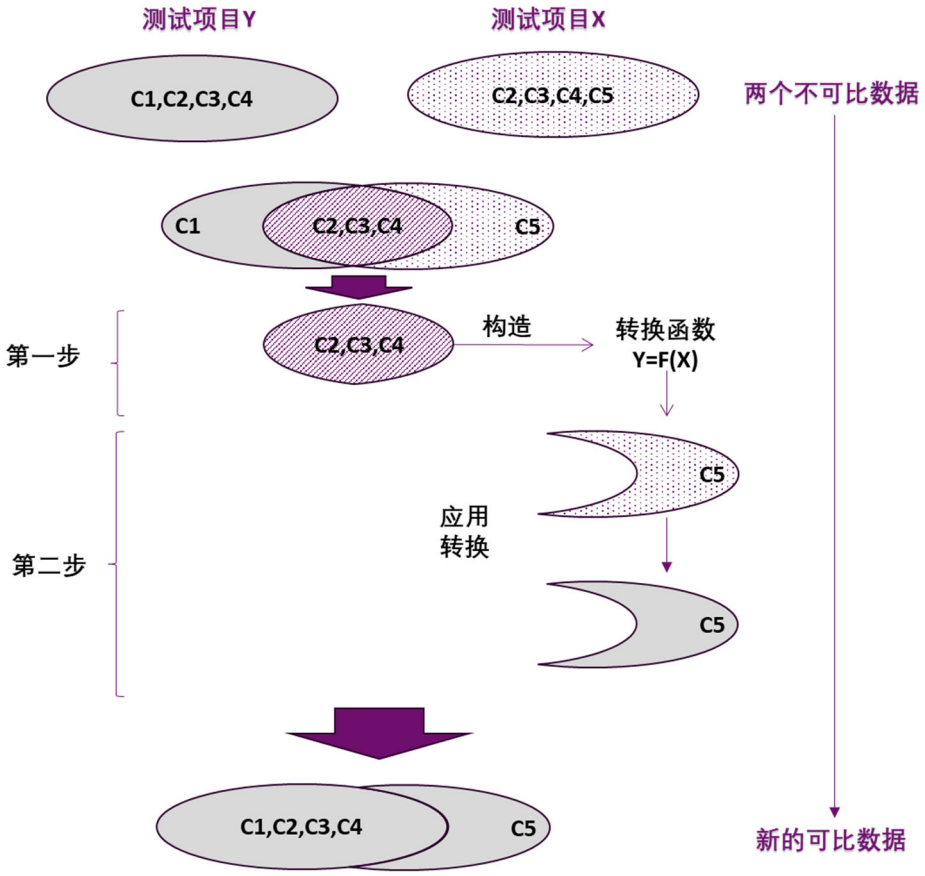

注：作者自绘。

**图 4-2 不同测试相同之间构建转换关系的方法框架**

此外，不管何种目的，如前所述，该类研究通常分为两个步骤：第一步，利用锚点构建函数（工具），第二步，应用该函数（工具）实现转换。如下图，以相同国家（见 4-2-2）扩大国家样本目的为例，假设有两个测试项目 X 和 Y（图中以不同的底纹阴影表示），都是针对学生数学素养开展测试，但覆盖不同

[26] 在本文中，共同、重叠、一样、重复代表相同的含义。

的国家和地区，其中测试项目 X 覆盖了国家 C1、C2、C3、C4，测试项目 Y 覆盖了 C2、C3、C4、C4。其中 C2、C3、C4 同时参加两种测试（如图中交叉的阴影部分），而 C1 和 C2 都参加一种测试。为扩大国家样本，我们首先要通过同时参加 X 和 Y 两种测试的国家样本（C2、C3、C4）构造出表现两种测试得分之间数量关系的转换函数，再通过该转换函数，将只参加测试项目 Y 的国家得分转换为测试 X 项目的得分，或将只参加测试项目 X 的国家得分转换为测试项目 Y 的得分。以 X 转成 Y 为例，我们将得到五个国家（C1、C2、C3、C4、C5）可比的数据。




注：以 X 转成 Y 为例。

**图 4-3 研究步骤示意图（以相同国家为例）**

**1. 以相同个体为锚点**

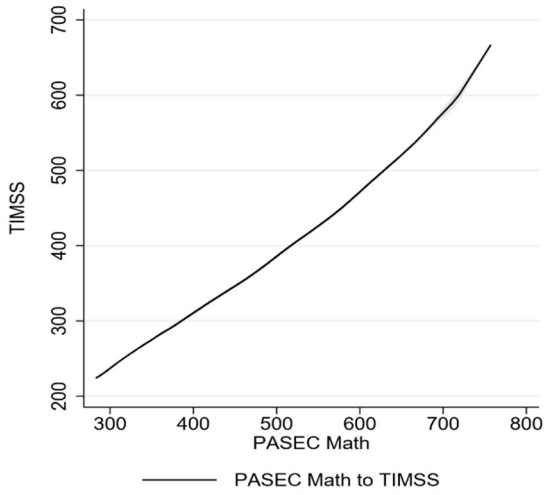
当两个测试项目由相同的学生群体参加时，并且测试学生相同的能力时，由于学生能力的稳定，因此测试项目之间的不同将直接反映在相同学生在两个测试项目中的得分不同上。此时，相同的学生群体作为锚点，通过学生在两个测试项目上的得分使用局部线性回归，直接构建两个不同测试项目得分之间的

映射函数。用公式表示为：

$$ScoreDistribution_Y \leftarrow ScoreDistribution_X \quad (4\text{-}3)$$

该映射关系表示测试项目$X$的每一个分数都对应着测试项目$Y$的一个分数。在得到映射关系后，便可以通过用于扩展国家样本等目的。比如，有一个国家学生只参加测试项目 X，未参加测试项目 Y，我们就可以根据 X→Y 的得分对应表，得到这个国家每个学生如果参加测试项目 Y 的得分。

举例而言，Patel & Sandefur (2020) 在印度召集了一组学生，让他们同时接受 PASEC 和 TIMSS 测试，然后运用局部线性回归对这些学生的 PASEC 和 TIMSS 测试得分数据进行估计，构建起这两种测试得分之间的映射关系。如图中的曲线，它就是一条反映 PASEC 和 TIMSS 测试得分之间映射关系的曲线，在这条曲线上，PASEC 与 TIMSS 测试得分存在一一对应关系。




注：来源于 Patel & Sandefur（2020）。

**图 4-4 以两维曲线表示的映射关系**

在得到映射关系后，便可以通过用于扩展国家样本等目的。譬如，有一个国家$c$的学生只参加 PASEC 测试，未参加 TIMSS 测试，我们可利用 Patel & Sandefur (2020) 所构造的得分对应表，将国家$c$的学生的成绩转换为 TIMSS 测试的得分，反之亦然。

以个体作为锚点的益处在于，它是对学生个体微观数据进行转换，因此转换后得到的数据依然是学生个体层面的微观数据，而非国家或地区层面的加总数据，研究者可以利用学生微观数据做更加细致的分析，实现国家或地区宏观层面无法达成的研究意图。并且在得到映射关系后，由于国际性和地区型测试

项目的稳定，即各年份可比，因此这一映射关系可以用于后来的测试。

不过，这种方法也有一些缺陷。首先，以一国或少数几个国家学生作为锚点所获得的不同测试得分的映射关系能否适用于其他国家学生，这是成疑的，因为该组学生可能更擅长某一测试而不擅长另一测试，使得得到的分数对应表存在偏差。其次，以个体作为锚点来构造映射关系，通常需要研究者亲手实施一项测试，所消耗的人力、物力和财力较大，总体成本较高。第三，映射关系应当可以实现任意分数的转换，尤其是高分和低分区间，因此需要保证各分数段上都有足够多的学生用以拟合映射关系。第四，除此以外，想要保证映射关系的有效性，在测试的实施全过程，如测试试题组成、学生选择、考试实施等过程，都需要严格保证规范和科学可信，这无疑是较难达到的。目前文献中只有 Patel & Sandefur (2020) 采用此种方法。

**2. 以相同总体（国家）为锚点**

我们首先要区分两个概念，相同总体和相同国家。当两个测试项目的参与国家都相同时（虽然因抽样问题，学生并不是同样的学生），此时属于相同总体。然而在现实中情况常常是只有一部分国家同时参与两个测试项目，此时研究常将这些同时参与两个测试的部分国家作为锚点（为区别我们称之为相同国家）。

为了便于理解这两种情况的不同，我们假设有两个测试项目 X 和 Y。当测试项目 X 和 Y 都测量了相同国家 C1、C2、C3，此时以三个国家形成的总体作为锚点，是相同总体；当测试项目 X 测量了国家 C1 和 C2，测试项目 Y 测量了国家 C2 和 C3，此时 C2 作为锚点，是相同国家。

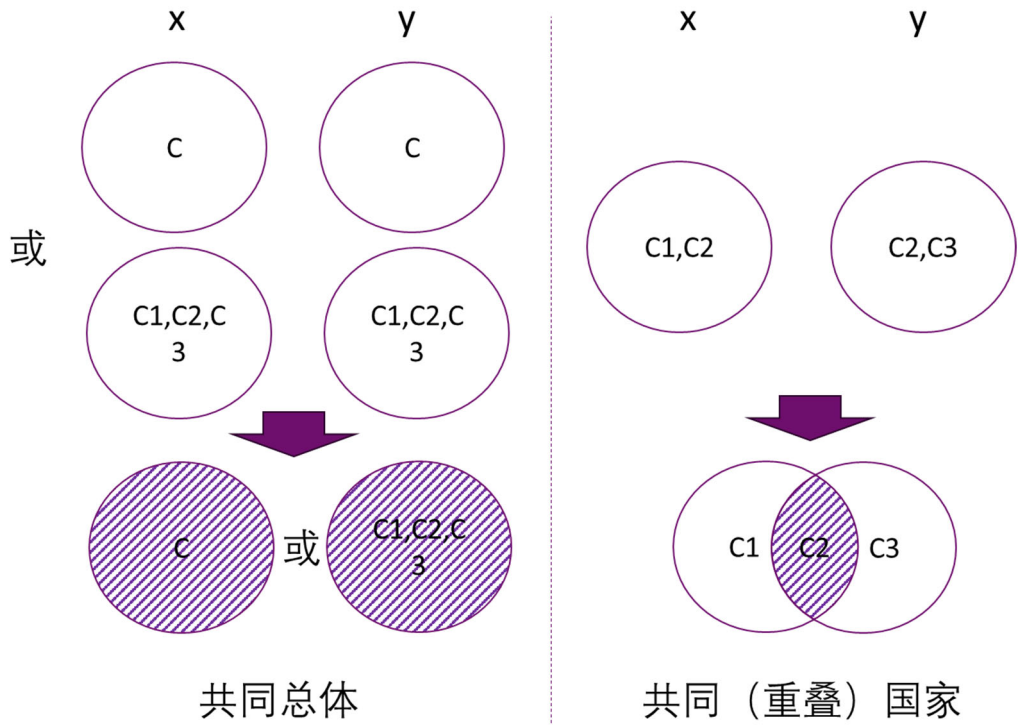



**图 4-5 共同总体和共同国家区分示意图**

两种情况其实反映了两种不同目的。对于相同总体，其目的在于整合两个测试各有的信息。譬如，测试 X 是统一实施的，C1、C2、C3 三个国家的成绩是可比的，但缺乏国家内部各省（州）的信息；测试 Y 是各国国家各自实施的，C1、C2、C3 三个国家成绩是不可比的，但各国实施测试时区分了国家内部各省（州），拥有的各省（州）的可比信息，通过测试 X 和测试 Y 的三个国家各自作为相同总体，整合两个测试各有的信息，便可以实现省（州）的跨国比较。

对于相同国家，其目的在于扩展样本国家数量。譬如，通过 C2 国家为锚点国家，我们可以利用确定两个测试项目 X 和 Y 之间的转换函数关系，并利用这一关系，预测出只参加测试项目 X 的国家 C1 如果参加 Y 测试项目的得分，以及只参加测试项目 Y 的国家 C3 如果参加 X 测试项目的得分，以此我们就可从只拥有两个国家的可比数据扩展到拥有三个国家的可比数据。

相同总体（国家）的方法需要基于两个前提假设：一是总体代表性假设，即每个国家被抽中参与两个测试项目的学生都应该能够代表相同的总体（Linked tests must test the same underlying population）[27]；二是内容相同假设，

[27] 事实上，由于不同技能测试项目的测量对象年龄或年级并不相同，使得该假设难以成立，为使得该问题

即两个测试项目所测量的技能是相同或相似的（Tests should measure similar proficiencies），譬如虽然在设计上有所差别，但 PISA 和 TIMMS 测试测量了学生的数学素养[28]。

在实际研究中，基于相同总体（国家）的方法数量最多。这些方法按照核心假设主要分为两种：

第一种的核心假设是微观分布相似假设。这是有关两个不同测试项目中相同总体（国家）的学生分数分布的假设，其认为如果不同测试项目都能够准确衡量各国总体学生的认知技能分布，那么相同总体（国家）的所有学生在不同测试项目上的得分分布形状将是一样的[29]，只在分布的均值和标准差上有所差异。由于这一种假设基于微观个体组成的分布，因此可以获得学生个体微观的数据[30]。基于这一假设的方法主要包括均值标准差转换和同分位点连接（Equipercentile Linking），其中后一种是相同总体（国家）方法中唯一的非参数连接方法。

第二种的核心假设是测试项目差异假设。这是有关相同国家的分数（均值）差异的假设，其认为相同国家中的各国国家在两个测试项目上得分的系统性差别应该来自两个测试项目的不同，而非国家的不同。与第一种假设不同的是，基于第二种假设的方法强调的是总体的差异，因此使用的国家均值来构造转换函数，因此这一方法获得是国家层面的总体数据，无法获得个体微观的数据。并且此类研究都是为用两个测试的重复国家为锚点，研究的目的都在于扩展样本国家。基于这一假设的方法主要包括线性连接（Line Linking）以及对此加以限定的均值连接（Mean Linking）和比率连接（Ratio Linking）。

但不管是哪一种假设，基于相同国家也都是在使用重叠国家的情况估计总体的情况。

---

产生的误差较小，在构建转换函数过程中，因尽量使用相近的群体，如使用 TIMSS 八年级和 PISA(15 岁）的数据来构建转换函数。

[28] 此处所表达的是，虽然 PISA 和 TIMSS 都数学数据，但是题目的不同使得他们测量偏向不同，数学下面也是划分很多小项的，他们的小项可能不一样，比如一个可能更偏向代数计算能力，一个更偏向空间几何能力。

[29] 由于测试项目都是根据项目反应理论测的，所以更具体而言，应该是两个测试项目得分分布都是正太分布。然而所有的方法其实都只需要两个测试分布是一样的，不需要一定都是正太的。

[30] 需要注意的是，在 Hanushek 早期的一系列研究中（Hanushek & Kimko, 2000、Hanushek & Woessmann, 2012a、Hanushek & Woessmann, 2012b），其虽然同样基于微观分布相似假设的均值-标准差转换思想，但其使用的是国家的总体均值，而非学生个体微观数据。并且在使用这一方法的非 Hanushek 研究中，也使用的是国家的总体均值，如 Altinok et al. (2018)。实际上，这些研究时间上是在将每个国家当成一个个体，使用这些国家个体形成的分布。在 Hanushek 近期研究中，其才将此方法应用于微观个体数据，如 Hanushek & Woessmann (2015)、Gust et al. (2024)。

如下表所示，在本文中，这一部分所有的方法几乎都已经被用来扩展样本国家数量，但目前只有均值标准差方法被用来整合两个测试各有的信息。

**表 4-1 目的-方法-假设对应表**

| 相同总体：整合信息 | | |
|---|---|---|
| 相同国家：扩展数量 | 均值标准差转换 | 微观分布相似 |
| | 同分位点连接 | |
| | 线性转换 | 测试项目差异 |
| | 均值连接 | |
| | 比率连接 | |

此外，需要注意的是，本文中对方法的命名与其他文献有些许差别，下表给出了方法命名的对比。

**表 4-2 方法名称在本文中与其他文献中的对应表**

| 本文中的方法命名 | | 其他文献中的命名 |
|---|---|---|
| 微观分布相似 | 均值标准差转换 | 线性连接（Altinok et al., 2018；Angrist et al., 2021） |
| | 同分位点连接 | 同分位点连接（Altinok et al., 2018） |
| 测试项目差异 | 线性转换 | 回归（Angrist et al., 2021） |
| | 均值连接 | 均值连接（Altinok et al., 2018） |
| | 比率连接 | 伪线性连接（Altinok et al., 2018）；比率连接（Patrinos & Angrist, 2018） |

注：作者自制。

（1）微观分布相似假设：均值标准差转换

在第一种微观分布相似假设下，由于两个测试项目的分布的形状是一样的，只在分布的均值和标准差上有所差异，因此只需要利用分布之间的转换公式，调整两个分布的均值和标准差[31]，便可以构造如下转换函数：

$$Score_{Yi} = \frac{(Score_{Xi} - \mu_X)}{\sigma_X}\sigma_Y + \mu_Y \quad (4\text{-}4)$$

其中$Score_{Yi}$和$Score_{Xi}$分别是个体$i$在测试项目$X$和$Y$中得分[32]；$\mu_X$和$\sigma_X$为锚

[31] 这一公式类似于对一个分布进行标准化后再逆标准化，而标准化只是对分布的尺度进行调整，并不会改变数据的原始信息。

[32] 对公式进行变形可以获得：$Score_{Yi} = \frac{\sigma_Y}{\sigma_X} * Score_{Xi} + \mu_Y - \frac{\mu_X}{\sigma_X}\mu_Y$，因此该方法也被 Altinok et al. (2018)成为线性转换。由于此方法实际使用的是微观分布相似假设，因此我们称之为均值标砖差转换，而非线性转换。即使当此方法应用于宏观国家层面数据，如 Hanusheck 的一系列研究，但其还是基于的是分布，只不过是从微观个体的分布变为宏观个体的分析。

点国家学生在测试项目 X 中的得分均值和标准差，$\mu_Y$和$\sigma_Y$为锚点国家学生在测试项目 Y 中的得分均值和标准差。

目前，能够完全算的上相同总体的只有 Reardon et al. (2021)，其意图是比较各州之间学区的得分，但是全国统一的 NAEP 测试项目虽然各州之间可比，但是却没有学区信息；同时由各州组织的测试项目虽然各州之间不可比，但却拥有洲内各学区的成绩信息。此时，相同总体为同时参加全国测试项目的也参加了自己的测试项目的各个州（State），通过整合两个测试的信息，可以实现学区的跨州比较。

在大多数研究中，通常是相同国家：现实中，情况常常是只有一部分国家同时参与两个测试项目（此时，我们称之为相同国家），比如在国际学生测试项目中，通常一些国家同时参与了 TIMSS；一些国家只参与了 TIMSS，但没有参与 PISA；一些国家只参与了 PISA，没有参与 TIMSS。在缺少严格相同总体的情况下，研究此时只能够用两个测试项目中尽可能多的相同国家$C$作为锚点，用这些相同国家的所有学生得分分布的均值和标准差来推测总体学生分布的均值和标准差（Gust et al., 2024），并改造转换函数：

$$Score_{Yi} = \frac{\left(Score_{Xi} - \mu_X^{pooled}\right)}{\sigma_X^{pooled}} \sigma_Y^{pooled} + \mu_Y^{pooled} \qquad （4\text{-}5）$$

其中$\mu_X^{pooled}$、$\sigma_X^{pooled}$、$\mu_Y^{pooled}$、$\sigma_Y^{pooled}$为由重复国家中学生个体组成的样本在两个测试项目的均值和标准差；$Score_{PISA,i}$和$Score_{TIMSS,i}$分别为学生$i$在 PISA 的得分，以及对应在 TIMSS 中能够获得的得分。在程序上，此类研究通常分为两步，第一步利用重复国家构造转换函数，第二步利用转换函数将只参与某一测试项目的国家的学生分数转换为目标测试的分数。

从公式中可以看出，以上转换事实上由两重转换组成：由均值的水平转换（或水平调整）和由标准差的差异转换（或差异调整），这两重转换分别对应了项目反应理论（IRT）[33]中的难度（Difficulty）和区分度（Discrimination）。水平转换解决的是由于两种测试项目难度不同，导致的相同的人在不同测试项目中得分不同的问题，并且这里假设难度导致的差异是对所有个体都是一样的（当标准差一致时，这个差异为$\mu_X - \mu_Y$）；差异转换解决的是由于两种测试项目区分度不同，导致的两个相同的人在不同测试项目中得分差异不同的问题，

[33] 项目反应理论内容见相同题项部分。

比如在一个测试项目中相差 10 分的两个人在另一个测试项目中只相差了 2 分。

由于两个测试项目的分布形状一致，通过转换，学生在总体的位置将保持不变。然而由于两个测试项目在内容上会存在一定的差异，这使得内容相同假设并不会严格满足，如 TIMSS 关注的是学校课程内容，PISA 关注的是现实问题（Hanushek & Woessmann, 2012a），体现在测试题目上存在不同，使得在内涵上这两个测试的数学素养存在一定差异。在这种情况下，两个测试项目的学生成绩分布形状会存在一定差异。这时候使用分布转换后，学生个体在分布中的位置会发生轻微变化，这一点时常无法避免，但各测试之间（如 TIMSS 的数学测试和 PISA 的数学测试分数之间的）较大的相关性使得内容相同假设不严格满足产生的误差在可接受范围内。因此，虽然这一方法可以获得学生个体微观的数据，但研究通常将关注点转移至分布均值而非某个学生的得分来降低这一误差的影响；如果使用转换后的学生个体微观分布的其他信息应当保有警惕。

此外，在严格数据的限制下，可能同时参加两个测试项目只有一个相同国家，此时只能够使用一个国家当作“瞄点”进行转换（Angrist et al., 2021）。均值与标准差转换的实质是用锚点国家学生个体的分布来推测潜在总体的分布，很明显，用单个国家的学生样本去推测总体的学生样本得分的分布无疑可信程度较低，因此用一个国家作为锚点进行转换，结果无疑会存在较大的误差[34]；实际上，如果用于连接的国家数目较少，使用这一方法转换结果的可信度便会因此而变得可疑（Gust et al. 2024）。

（2）微观分布相似假设：同分位点连接（Equipercentile Linking）

在学生层面，同分位点连接（Equipercentile Linking）是一种常见的没有使用项目反应理论的、非参的测试项目分数比较方法（Kolen & Brennan, 2014），这一方法由 Braun & Holland (1982)开发。

在微观分布相似假设下，相同总体（国家）的所有学生在不同测试项目上的得分分布形状将是一样的，那么可以直接利用同分位点等值将两个分布连接起来。

该方法通常包含两个要素：

一是同分位数点数值。即根据分数分布的分位点，将对应的分数连接起来，

[34] 由于两个测试的内容的不同，有可能某一国家的某些学生更擅长 X 测试，另一些学生更擅长 Y 测试，因而使得同一国家的两个测试分布形状不一致，这时候，只有利用更多国家的大量个体，两个分布将更接近形状相似的正太分布。

较为直观的如下图所示：

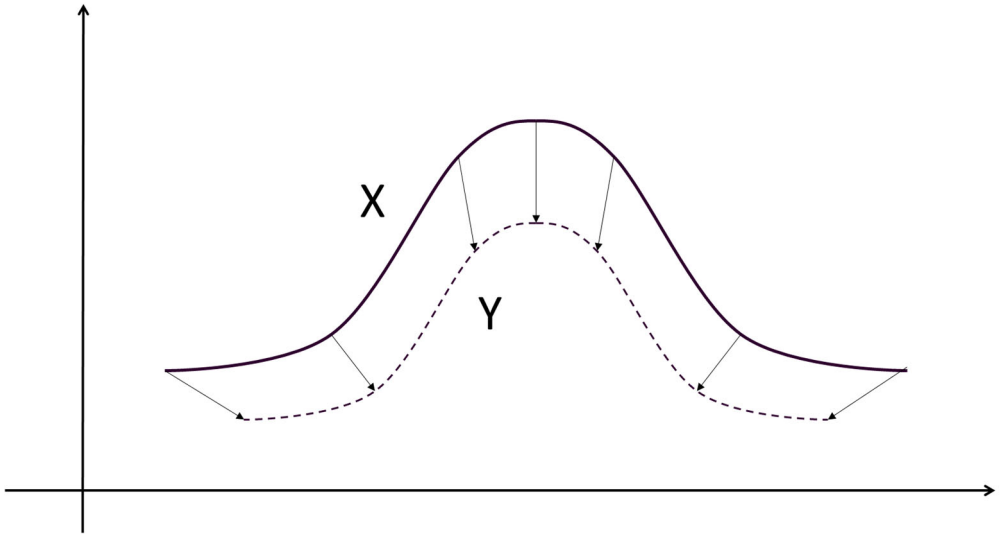



**图 4-6 同分位点连接示意图**

图中有两个成绩分布 X 和 Y，X 的没一个分位点（12th 百分位点）的取值（即分数）都与 Y 的同分位点（12th 百分位点）的取值对应。若用公式表达，可以写成下式：

$$Q_X^q = Q_Y^q \quad （4\text{-}6）$$

二是平滑处理。在同分位点连接中，因为分数较为离散，我们可能无法找到某个分数或者某个分位点分数对应的另一个分布的分数和分位点分数，因此需要平滑处理。比如，X 的 100 分和 105 分分别与 Y 的 103 分和 110 分对应，但因为分数取值比较离散，因此不知道 X 的 102 分和 Y 的多少分对应；又比如，在理想状态下，X 的 47th 百分位点应该对应着 Y 的 47th 百分位点的分数，但由于分数取值离散，使得在实际中最接近的分数可能分别对应着 X 的 47.2th 百分位点和 Y 的 47.6th 百分位点，这时候虽然我们可以获得粗略的匹配对应，但并不足够精确。

在这种情况下，我们虽然可以利用百分位排名（Percentile Ranks），但精度不足；也可以增大样本量来缓解这一问题，但往往不足。为此，平滑方法被开发出来处理抽样误差，获得能够最接近潜在总体的实证分布和分位点连接关系（Altinok et al., 2018）。

根据两种平滑处理方式，同分位点连接可以分为以下两种：前平滑同分位点连接（Presmoothed Equipercentile Linking）和后平滑同分位点连接（Postsmoothing Equipercentile Linking）。在前平滑同分位点连接（先平滑后等

值）中，通常先采用多项式对数线性（Polynomial Loglinear）对得分分布进行平滑处理，然而在分位点取值构建对应表（Holland and Thayer, 2000）；在后平滑同分位点连接（先等值后平滑）中，先利用得分分布进行同分位等值，后采用三次样条（Cubic-spline）进行平滑处理（Kolen, 1984）。

事实上，由于是基于分位点，因此该方法并不严格要求两个测试项目的分布形状一致，但要求相同的个体在两个测试项目中所处的分布位置是一样的。这相当于放宽了假设。所以，这一方法最适合应用于两个测试项目的难度呈非线性变化（Altinok et al.，2018）。

和之前一样，目标通常是基于同时参与两个测试项目的国家将只参与了某一个测试项目的国家的学生个人分数转换成另一个测试项目的分数。因此这一方法的基本途径是，用相同国家的学生得分分布推测总体的得分分布，然后构建两个测试项目分布之间对应关系，最后对只参与了某一测试项目的国家学生层面数据进行转换。并且，同样，相同国家的数目是影响连接质量的重要因素。因为只有相同国家数目足够多，学生样本才足够多，两个测试的分布才能接近于总体的分布；不同分位点也才有足够多的样本来估计。使用这一方法的文献主要为 Altinok et al., (2018)、Sandefur (2018)。

（3）测试项目差异假设：线性转换（Liner Transforming）

在相同国家中，各国成绩在两个测试项目分析的系统性差别应该来自两个测试项目的不同，而非国家的不同。因此，可以直接通过使用 OLS 估计下式获得转换参数：

$$\mu_{Y,c} = \alpha + \beta * \mu_{X,c} + \varepsilon_c \quad （4\text{-}7）$$

其中，$\mu_{X,c}$和$\mu_{Y,c}$分别为锚点国家$c$在测试$X$和测试$Y$中的得分均值，$\alpha$和$\beta$用于捕捉两个测试项目之间的系统性差异[35]。和均值标准差转换一样，$\alpha$和$\beta$分别近似对应着分布转换中的水平调整和差异调整，反映两个测试项目的难度和区分度差异。模型中的$c$表示在同一轮测试项目中，那些同时参与两个不同测试项目的国家，这一方法同样是使用共同国家来估计总体。在利用相同国家构造转

[35] 该方法与 Altinok et al.（2018）中的线性转换并不一致，因为该方法的 β 估计系数为$\frac{cov(X,Y)}{var(X)}$，而在 Altinok et al.（2018）中为$\frac{\sigma(Y)}{\sigma(X)}$，两者的关系为：$\beta = \rho * \frac{\sigma(Y)}{\sigma(X)}$，其中$\rho$为相关系数。但不可否认的是，从均值标准差转换也可以推导出均值连接和比率连接，这也正是 Altinok et al.（2018）中的行文顺序。

换函数后，便可以将只参与 X 测试的国家的总分转换为 Y 测试下的总分。

需要注意的是，在 20 世纪 90 年代后，各测试相同通过相同题项（见下面介绍）实现了轮次之间的可比，这也意味着两个测试项目之间的差异将在各轮之间保持不变。因此，上式可以使用多轮数据来估计（$r$表示轮次）：

$$\mu_{Y,c,r} = \alpha + \beta * \mu_{X,c,r} + \varepsilon_{c,r} \quad （4\text{-}8）$$

用回归估计转换参数这一方法最近才被使用（Angrist et al.，2021），该方法虽然同样受限于共同国家的样本数量，但会随着时间的推移、测试项目的持续开展而出现越来越多的可用于估计的样本，估计的转换系数准确性也将越来越高。但这方法存在的问题同样在于，每多一轮测试，方法估计的参数都将会有细微的不同。

（4）测试项目差异假设：均值连接（Mean Linking）

如果同一个国家参与了两个测试项目（$X$，$Y$），获得了两个分数（$Score_X$和$Score_Y$），这两个分数的关系可以由下式给出：

$$Score_Y = a + Score_X \quad （4\text{-}9）$$

基于这一想法，均值连接认为两个测试项目之间的差别都是一个固定的常数：$\alpha$。也就是说，两个测试之间的差异将由下式给出：

$$Score_Y = \alpha + Score_X + \varepsilon_c \quad （4\text{-}10）$$

常数$\alpha$可以使用同时参与两个测试项目的相同国家来估计[36]：

$$\alpha = \mu(Y) - \mu(X) \quad （4\text{-}11）$$

其中，$\mu(Y)$和$\mu(X)$为相同国家在两个测试项目中的均值平均值。由于视野局限，作者尚未发现将此方法用于 20 世纪 90 年代后的测试项目的文献，但此方法和线性转换一样，当用于 20 世纪 90 年代后的测试项目的转换时，可以使用多年数据来估计转换系数。

与线性转换相比，这一方法通过限定转换函数为简单加减形式，这其实是限定了两个测试项目之间差异的形式，如果实际测试项目差异的形式并非如此，则会产生较大的误差。只有当两个测试项目的区分度一致时，这一方法的转换

[36]也可以利用回归，限定斜率为 1，直接进行估计。

结果才和线性转换一致。

在 Altinok et al. (2018)中首次对这一方法进行了介绍，不过早在 Hanushek & Kimko (2000) 一文中已有使用。

（5）测试项目差异假设：比率连接（Ratio Linking）

这一方法主要来自世界银行的一系列文摘（Altinok & Murseli，2007；Altinok et al., 2014；Altinok et al., 2018；Patrinos & Angrist, 2018）。如果同一个国家参与了两个测试项目（$X$，$Y$），获得了两个分数（$Score_X$和$Score_Y$），这两个分数的关系可以由下式给出：

$$Score_Y = \beta * Score_X \quad （4\text{-}12）$$

基于这一想法，比率连接认为两个测试之间的差异是一个固定的比率：$\beta$。也就是说，两个测试之间的差异将由下式给出：

$$Score_{Y,c} = \beta * Score_{X,c} + \varepsilon_c \quad （4\text{-}13）$$

$\beta$可以使用同时参与两个国家的相同国家来估计[37]：

$$\beta = \frac{\mu(Y)}{\mu(X)} \quad （4\text{-}14）$$

其中，$\mu(Y)$和$\mu(X)$为相同国家在两个测试项目中的均值平均值。此方法和线性转换一样，当用于 20 世纪 90 年代后的测试项目的转换时，可以使用多年数据来估计转换系数。

与线性转换相比，这一方法通过限定转换函数为简单乘积，这其实是限定了两个测试项目之间差异的形式，如果实际测试项目差异的形式并非如此，则会产生较大的误差。只有当两个测试项目的难度一致时，这一方法的转换结果才和线性转换一致。

（6）小结

没有任何的方法所转换的得分是完美的，重要的是，这些类似某一国家学习进度不同非测试项目问题所产生的影响足够小，在文章中所列举的方法都采用了一定方法来减小这些因素的影响，比如“均值与标准差转换”“同分位点链接”中基于锚点国家学生个体样本 Pool 后计算转换参数和链接；线性转换中残差项，均值连接法（Mean Linking）和比率连接法（Ratio Linking）中使用锚点

[37] 也可以限定截距项为 0，使用回归直接进行估计。

国家在两个测试项目中的均值平均值$\mu(Y)$和$\mu(X)$。

**3. 以相同题项为锚点：项目反应理论连接（IRT Linking）**

可以想象，如果两个测试项目所包含的题项完全相同，那么即使对于不同的学生群体，其得分也是直接可比的。然而，这一理想并不切合实际，即使是相同测试，在不同轮中也是有所差别的。事实上，如果我们放低要求，只需要在两个测试项目中有一定量的重叠题项，就可以以这些重复题项作为锚点，依据一定方法实现两个测试项目得分的可比。

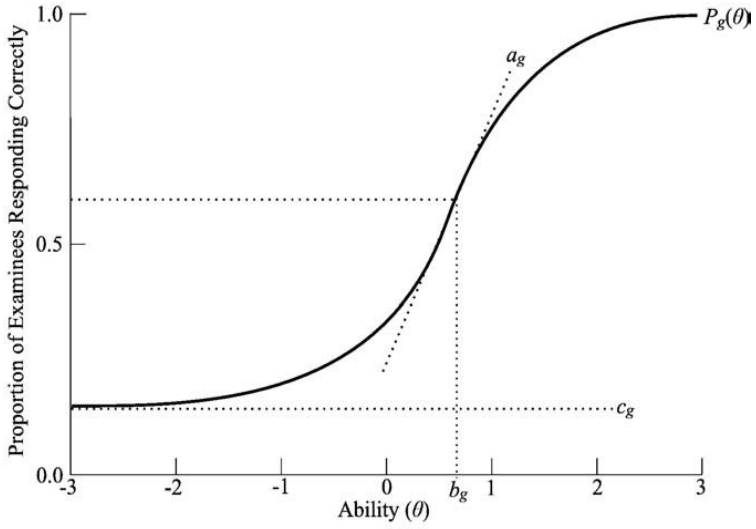
通过相同题项可比的基础在于两个测试项目都依赖于项目反应理论（Item Response Theory，IRT）。项目反应理论作为现代心理测量理论，被广泛地应用于国际性和地区性的测试项目。项目反应理论认为学生正确回答给定测试题项的概率为学生特征和测试题项特征的函数。

具体而言，在一个最常用的三参数逻辑模型（3PL）中，对于一个二值型答案题目（$X_{ig} \in \{0,1\}$，其中 0 表示回答错误，1 表示回答正确）而言，项目反应函数（Item Response Function，IRF）给出了一个素养为$\theta_i$的个体$i$正确回答难度（Difficulty）为$b_g$、区分度（Discrimination）为$a_g$、正确猜测可能性（Probability of Guessing Correctly，又称猜测度）为$c_g$的题目$g$的可能性[38]：

$$P\left(\left(X_{ig}=1\right)\middle|\theta_i, a_g, b_g, c_g\right) = c_g + \left(1-c_g\right)\frac{EXP\left(a_g*\left(\theta_i-b_g\right)\right)}{1+EXP\left(a_g*\left(\theta_i-b_g\right)\right)} \quad (4\text{-}15)$$

其中$\theta$又被称作潜变量（Latent Variable），通常为各种各样的思维、能力、素养、特征等，在国际性和地区性学生测试中，通常为数学素养、阅读素养和科学素养。下图是三参数模型的项目反应曲线（IRC），给出了参数作用的直观呈现。

[38] 使用 IRT 的测试项目的分数主要由难度和区分度，这也决定了测试之间的转换连接应当对难度和区分度进行调整。




注：引用自 Das & Zajonc（2008）。

**图 4-7 项目反应曲线**

根据参数的不同，上式可缩简为单参数模型（难度）和双参数模型（难度、区分度）。在知道学生正确回答与否和题项特征参数的情况下，根据模型估计出学生特征参数（即学生的能力、素养），获得学生测试项目得分。项目反应理论有两个关键假设：所测量潜变量的单维性假设和参数不变性假设。单维性假设是指，组成某个测验的所有项目都是测量同一潜在特质；参数不变性假设是指题目特征参数对于任何人群都是固定的，不受考生能力分布的影响。

原则上，如果两个测试项目拥有一定量的相同题项，便可以依据相同题项连接两个测试项目，实现两个测试项目得分的可比。假如，我们以实施了三轮且每轮测试都只有两道题的测试项目为例，每轮的题目为（题 1，题 2）（题 2，题 3）（题 3，题 4），依次的两轮都有一个相同的题项：（题 2）和（题 3）。当题 1 的特征参数固定已知时，由于题 1 和题 2 都应当估计出相同的能力（潜变量），因此就可以确定题 2 的特征参数。同样在给定题 2 的特征参数的情况下，就可以估计题 3 的特征参数。依此类推，就可以将第三、二轮测试项目与第一轮测试项目关联起来，实现多轮测试项目的可比（Das & Zajonc，2010）。需要注意的是，虽然前面以一道重复题项举例，但在实际操作中，两个测试项目应当拥有一定比例的重复题项，才能达到减少链接误差的目的（Hastedt & Desa, 2015）。

事实上在 20 世纪 90 年代后，国际性和地区性测试项目都会在不同轮次的测试项目中设置一定量的重复的题项，以此实现各轮测试项目得分在时间上可

以直接比较（Angrist et al., 2021）[39]。

根据题项特征参数设置的不同，会产生不同的方法，我们在这里主要介绍以下两种构建连接的方式（Sandefur, 2018）：

（1）题项特征参数等同

在这种情况下，利用目标测试项目（一般为地区性测试项目，如 SACMEQ）中的相同题项，使用参考测试项目（一般为国际性测试项目，如 TIMSS）中这些相同题项的特征参数[40]。

也就是说：

$$a_{gr} = a_{gt} \quad （4\text{-}16）$$

$$b_{gr} = b_{gt} \quad （4\text{-}17）$$

$$c_{gr} = a_{gt} \quad （4\text{-}18）$$

其中下标$r$和$t$分别代表参考测试项目（Reference Tests）和目标测试项目（Target Tests）。

使用这一方法的文献包括上述的举例，以及 Das & Zajonc (2010)、Singh (2014)、Sandefur (2018)。

（2）题项难度和区别度等同（Mean-sigma Linking）

与题项特征参数等同不同的是，它不是将目标测试项目和参考测试项目的相同题项的特征参数一致，而是要求保证相同题项的平均难度和区分度在两个测试中保持不变[41]。基于项目反应理论核心假设中的不变性假设，两个等值测试项目得分都可以通过一个线性转换所关联：

$$\theta_t = A_{rt} * \theta_r + B_{rt} \quad （4\text{-}19）$$

与之前相同总体（国家）中的一致，系数 A 用于调整区分度，截距 B 用于调整难度。相似的变换可以应用至题项的特征参数上：

---

[39] 对 2000 年前测试分数进行转换的研究见 Hanushek & Woessmann (2012a)，这些研究的锚点是可跨时间进行比较的测试，通常为美国的 NAEP 测试。

[40] 该方法的有效性依靠项目功能差异（Differential Item Functioning，DIF）来检验。

[41] 文献中没有提到猜测度如何设置。

$$a_{gr} = a_{gt}/A_{rt} \quad (4\text{-}20)$$

$$b_{gr} = A_{rt}b_{gt} + B_{rt} \quad (4\text{-}21)$$

理论上，只需要一个相同题项，就可以获得$A_{rt}$和$B_{rt}$两个参数，然而由于测量误差和模型的不完美拟合问题致使这一途径实际行不通。事实上，研究者被迫在每一个题项所得到的不同的$A$和$B$之间做出选择。

或者采取一个更简单的方法，被称之为均值标准差（Mean-sigma）法，通过相同题项的不同的特征参数$b$的均值和标准差来获得$A$和$B$：

$$A_{rt} = \sigma(b_{gr})/\mu(b_{gt}) \quad (4\text{-}22)$$

$$B_{rt} = \mu(b_{gr}) - A_{rt}\mu(b_{gt}) \quad (4\text{-}23)$$

在获得转换系数后，即可以直接应用于学生的素养参数$\theta_i$，将目标测试项目得分转换为参考目标得分。也就是说，这一方法可以看作是从题项特征参数层面估计相同总体（国家）中线性转换公式的参数。使用这一方法的文献包括Sandefur (2018)。

## 五、融合受教育年限与学生认知技能：一种兼顾量与质的方法

### （一）构建既含有质量也含有数量的指标

在关注教育获得的数量向关注教育获得的质量过程中，学界并非完全否定教育数量的重要性，将平均受教育年限抛弃，而是认为教育获得的数量和质量同等重要。因此一些研究试图结合教育质量和数量，出现了两类途径：线性结合和乘积结合。

在进入分析之前，我们需要回到一个问题：平均受教育年限只能衡量教育获得的数量，这一点经过众多学者研究的努力，已经成为共识。然而，学生认知技能是否能够衡量教育获得的质量？

根据人力资本理论和教育生产函数理论，认知技能是人力资本的重要组成部分，是学校教育、家庭教育、个人天赋等投入因素的共同产出，对于学校教育，现有研究又通常将简单分解为两个方面：教育数量和教育质量。也就是说，从学理上看，受教育年限指向的是个体所接受的学校教育的数量，它是一种教育投入指标；而学生认知技能是一种教育产出指标，它是教育数量和教育质量

的共同产出，而非直接的教育质量度量。

如此看来，受教育年限和学生认知技能是不同维度的两种指标，其不能简单地将它们融合在一起。

根据以上讨论，正确的做法应当是利用教育生产函数，从学生认知技能中分离出教育质量。同时，当学生处于同一年级或年龄时（即教育数量相同），学生认知技能只由教育质量所决定，学生认知技能可以用来反映教育质量。也就是说，只有当学生处于同一年级或年龄时，我们才能使用学生认知技能来度量教育质量；当学生不存于同一年级或年龄时，则需要利用教育生产函数，从学生认知技能中估计出教育质量。[42]我们将在第八章重新回到这一点上。

以上分析也告诉读者，学生认知技能的使用的关键并不仅仅是各测试项目的可比，还要求学生的一致。然而，现有研究在使用学生认知技能时，通常是“数据有无优先，学生一致靠后”：即用其他测试数据来填补空缺，不考虑或者强制忽略学生的年龄和年级不一致（Gust et al, 2024）。

也有研究，根据成人认知技能估计不同时期成人所接受教育的质量，并结合成人的受教育年限，计算经质量调整后的受教育年限（Hanushek & Zhang, 2009）。不同于学生认知技能，使用成人认知技能，必须要利用教育生产函数从成人认知技能中估计教育质量。

下面我们就从人力资本（认知技能）生产函数出发，对这一方面的内容做详细介绍。

需要注意的是，以下谈论大部分都假定学校质量不能随着时间和年级而变化，也就是，今年的教育质量和明天的是一样的；3 年级的教育质量和 4 年级是一样的。

### （二）人力资本（认知技能）生产函数和线性结合与乘积结合

#### 1. 人力资本（认知技能）生产函数

根据教育生产函数理论，教育产出是学校教育、家庭教育、校外教育等因素的最终产物（Hanushek & Zhang, 2009）。其中学校教育对认知技能的影响是关于教育质量和教育数量的函数。结合 Hanushek & Woessmann (2012a)，人力资

[42] 当使用同一年级的认知技能作为教育质量时，由于此时的教育数量相同，因此研究并不会产生问题；当使用不同年级的认知技能作为教育质量时，由于此时的教育数量不想同，此时将学生认知技能作为教育质量会存在一定的误差。

本（认知技能）生产函数可以用下式表示：

$$H(CS) = \lambda F + f(n,q) + \eta A + \alpha Z + v \quad （5\text{-}1）^{43}$$

其中$H$为人力资本，在此处主要为学生/成人认知技能（CS）；$F$为家庭投入，$\phi(n,q)$为学校投入，包括学校教育的数量（$n$）和质量（$q$）[44]，$A$为个人天赋，$Z$为其他因素，包括劳动力市场经验、健康等等。

正如“Schooling is not learning”一样，“Learning is not just schooling”，然而在通常情况下，研究还是更加关注学校教育。因为与校外教育相比而言，学校教育对认知技能的影响更大（Filmer et al., 2020），因而其他部分通常会被一些研究所忽略：

$$H(CS) = f(n,q) \quad （5\text{-}2）$$

该生产函数传递的最为重要信息是，接受学校教育不代表就能一定促成个人技能的实质性发展与人力资本的有效累积。对于个人技能发展来说，学校教育的数量与质量只是投入，由学校教育投入到个人技能发展，需经历$f(Schooling, Quality)$的函数转换过程。

然而，$f(Schooling, Quality)$的函数形式是未知的，需人为设定。关于函数形式可以划分为两个点进行讨论，一是教育数量和教育质量的组合方式，而是教育生产效率，即教育数量和质量的组合如何影响教育产出。当前，关于教育数量和质量的组合方式，有两种形式：线性结合和乘积结合。

（1）线性结合

线性结合是指教育数量和教育质量的关系为线性组合，因此这一函数通常为：

$$H = f(rS + wQ) \quad （5\text{-}3）$$

在 Hanushek et al. (2017)、Angrist et al. (2020)等文献中，具体采用指数函数形式：

$$H = e^{rS+wQ} \quad （5\text{-}4）$$

其中，s 为教育数量，通常使用平均受教育年限；Q 为教育质量，通常采用

[43] 为了简单，上述里忽略了学科和时间的符号。
[44]此处指的是单位质量。

学生认知技能；$r$和$w$通常来自于微观的明瑟收入方程估计。

首先不难看出，在对两边取对数后，这一形式和明瑟收入方程十分接近，这可能也是采取对数形式的原因。其次，现有研究经常使用不同年级或年龄学生的认知技能直接作为教育质量放入回归会中，忽略了不同年龄或你那里学生的认知技能不完全等同于教育质量。

（2）乘积结合：质量调整后的受教育年限（Quality-adjusted years of schooling，QAYS）

乘积结合是指教育数量和教育质量在函数中是乘积形式：

$$H(CS) = f(SQ) \quad （5\text{-}5）$$

与之前不同的是，为了使这一乘积富有含义，$Q$在这里通常为质量调整系数，而非直接的教育质量，因此，$SQ$便可以理解为质量调整后的受教育年限[45]。正式的用公式可以如下表达：

$$QAYS_c = S_c * Q_c^b \quad （5\text{-}6）$$

其中$QAYS_c$为调整后的受教育年限，$S_c$为国家$c$的平均受教育年限，$Q_c^b$为质量调整系数。质量调整系数一般以某一个国家为基准，可以通过下式计算获得：

$$Q_c^b = \frac{q_c}{q_b} \quad （5\text{-}7）$$

其中，$q_c$为目标国家的教育质量，$q_b$为参考国家（Benchmark Country）的教育质量。

$QAYS$实际上是在不抛弃平均受教育年限的情况下，对平均受教育年限进行改造，使其包含质量维度的信息。

**2. 函数形式的选择：不同类型的认知技能生产函数**

对于线性结合，问题的关键在于估计出数量和质量两者结合的系数 r 和 w，如前所言，这主要是由微观数据估计。

而对于乘积结合而言，问题的关键在于估计出教育质量。教育质量反映了

[45] 这些研究包括 Hanushek & Zhang（2009）、Kaarsen（2014）、Filmer et al.（2020）、、Reiter et al.（2020）、Glawe & Wagner (2022)等。在一些研究中，质量调整后的受教育年限直接等同于人力资本，在另一些研究中，其还是会对其套一层函数来计算人力资本，如 Kaarsen（2014）。这里也可以看出，相关概念使用的混乱。

一个的学校教育系统生产力（Productivity；Filmer et al., 2020）或学校教育系统有效性（effectiveness；Kaarsen, 2014）。

教育质量的估计还需要回到人力资本（认知技能）生产函数中。

由于学生/成人认知技能通常是国家层面的均值，因此乘积形式的生产函数可以改写为：

$$CS_c = f(n_{c,l} * q_c) + p(X) + \varepsilon_c \quad （5\text{-}8）$$

$CS_c$为国家$c$的学生/成人认知技能，$n_{c,l}$为国家$c$的年级$l$所对应的教育年限，在现有的学生国际性测试项目中，$l$通常为 4 或者 8 年级；$q_c$为国家$c$的教育系统中不同年级的学生所接受的教育质量。因为假定教育质量不随年级而变化，因而教育质量的符号的下标只有国家$c$[46]。

即使教育数量和教育质量将以乘积的形式进入回归中，上述也需要明确生产函数的形式，才能估计出教育质量，这就需要对教育质量的含义做进一步的讨论。在现有的文献中，从投入产出角度有两种不同教育质量认识方式：

一是认为教育质量是一种投入，从投入到产出之间存在一个生产效率的问题。投入可以是边际收益递减，如在 Kaarsen（2014）中，$f(\cdot)$为对数形式：

$$CS_c = ln(n_{c,l} * q_c) + p(X) + \varepsilon_c \quad （5\text{-}9）$$

投入也可以是边际报酬不变，如在 Hanushek & Zhang (2009)，$f(\cdot)$为线性形式：

$$CS_c = \alpha * n_{c,l} * q_c + p(X) + \varepsilon_c \quad （5\text{-}10）$$

二是认为教育质量是一种产出，产出之间的累积等于总的产出。在本文的语境中，这意味着教育质量就是每年能够获得的认知技能得分。于是认知技能生产函数可以直接简化为[47]：

$$CS_c = n_{c,l} * q_c + p(X) + v \quad （5\text{-}11）$$

---

[46] 这一点也可以理解为是估计教育系统的平均质量。

[47]在这种情形下，如果教育质量可以随不同的年级而变化，生产函数就可以写为累加的形式：

$$CS_c = \sum_{j=1}^{n_{c,l}} q_{c,j} + p(X) + v$$

从不同的视角，选取合适的函数形式，施加特定的假设，即可以获得教育质量。

在利用认知技能生产函数估计出教育质量后，便可以利用公式构造质量调整系数，计算调整后的平均受教育年限。

**3. 计算 QAYS：一般步骤是分歧点**

因此这类研究的一般是顺序为利用学生认知技能数据，以及公式（26）或（27），估计出各个国家对应时期的教育质量 q；然后基于获得的各国教育质量，利用公式（29），计算教育质量调整系数$Quality^{benchmark}$；最后利于公式（28），利用对应时期的各国平均受教育年限数据获得 QAYS 数据。

在计算过程中，我们使用了某一时间点的学生认知技能数，如 PISA2018 年的 15 岁学生的得分，从学生认知技能数据中估计出来教育质量[48]。以及某一个时间点的而平均受教育年限数据，成人数据的时间点选择各异，有使用接近学生时间的刚毕业人口数据的如 2018 年时的 25-29 岁人口的平均受教育年限（Filmer et al., 2020），也有使用对应时期 25-64 岁人口的平均受教育年限数据（Kaarsen，2014；Altinok & Diebolt，2023）。成人平均受教育年限人口群体的选择，决定了 QAYS 衡量的是一国当前的教育系统，还是该国的人资本存量。但无论怎样选择，学生数据与成人数据都并不同期。然而从学理上讲，25-29 岁成人数据离学生数据时间上更近，其更有说服力，因此，QAYS 最适合的用途是衡量当前教育系统的质量。

换句话来讲，计算 QAYS 的前提在于拥有两点时间点的数据：一个时间点的学生认知技能数据和一个时间点的成人平均受教育年限数据。其中学生认知技能用于估计教育质量，然后对成人平均受教育年限进行调整。

从以上推演过程中，可以看出 QAYS 指标是对教育数量和质量的组合方式，以及人力资本生产函数形式施加了层层假设后才推导得到，这意味着研究的实施都依赖对认知技能生产函数的假定，而当前，学界对于人的认知技能生产的技术、方式等方面知识掌握还很不充分，这使得学者们在进行相关研究时不得不做出许多并未有太多经验证据支持的主观假设，因此 QAYS 指标构建存在着较大的测量偏误风险。认知技能生产函数“黑箱”是横亘在这一领域学者面前

[48] 理想状态下，我们应当使用相同时间点的教育数量和教育质量进行融合。理论上，我们应当使用 15 岁的教育数量，但这属于画蛇添足，毕竟学生认知技能已经是包含了教育数量和质量的指标。

的一座难以逾越的大山。

## 六、成人认知技能：一种最佳的人力资本测量指标

本文的讨论从平均受教育年限，到了学生认知技能，又到了教育数量和教育质量的结合。不由提出一个问题，以上的测量都是人力资本的最佳测量指标吗？不是的话，又会是哪一个指标？

本文认为在所有的指标中，最佳的人力资本测量指标是成人认知技能。

其原因有以下几点：其一，学生认知技能是流量指标，而成人认知技能数据是人力资本存量指标；这一方面是因为成人认知技能测试的参与者涵盖了劳动力市场上的各个年龄段，而学生认知技能通常只有某一年级或年龄，另一方面是因为学生认知技能数据是某一年级或年龄的，并非其接受完所有教育后的认知技能。

其二，根据教育生产函数，不论是教育数量（平均受教育年限），还是教育质量都属于投入，成人的认知技能最终的产出。第三，人力资本是凝结在人身上的知识、技能等要素，因此用教育数量和教育质量合成的人力资本测量指标（QAYS），显然也不如人力资本的直接测量结果好。

### （一）构建成人技能数据库：利用与学生数据的关联突破成人调查包含有限的国家的局限

现有的成人认知技能调查[49]包括国际成人文化素养调查（International Adult Literacy Survey，IALS）[50]、国际成人文化素养和生活技能调查（International Adult Literacy and Life Skills Survey，ALLS）[51]、国际成人能力评估调查（Programme for the International Assessment of Adult Competencies，PIAAC）[52]以及就业能力和生产能力调查（The Skills Towards Employability and Productivity (STEP) program）[53]。其中 PIAAC 包涵的国家数量最多，但也只提供 36 个国家

[49] 有关更多地成人认知技能调查的简要介绍，见 De La Fuente & Doménech (2024)、Reiter et al. (2020) 。

[50] 该调查由 OECD 在 1994-1998 年间实施，共调查了 22 个国家，调查内容包括散文阅读素养（Prose Literacy）、文件阅读素养（Document Literacy ）和数理阅读素养（ Quantitative Literacy）。

[51] 该调查由 OECD 在 2003-2007 年间实施，被当作是 IALS 的继承者，共调查了 11 个国家，在内容上，由算术素养（Numeracy）代替了之前的数理阅读素养（Quantitative Literacy），并新增了问题解决（Problem-solving）。

[52] 该调查由 OECD 在 2011 至 2018 年间实施，共调查了 37 个国家，在内容上，其由文学素养（Literacy）、算术素养（Numeracy）和技术丰富环境中问题解决能力（Problem-solving in Technology Rich Environments）组成。

[53] 该调查由世界银行（World Bank）在 2012-2017 年间在 17 个非 OECD 低收入和中等收入国家进

的数据。

可以说 PIAAC 参与国家的有限性极大的限制了其在宏观研究的使用。与此对应的是，随着学生调查测试项目的开展，越来越多的国家被包含在内。为此近来少数学者开始尝试将各国学生认知技能和成人认知技能数据联系起来，通过构建二者之间的转换关系来扩展成人认知技能数据的国家样本数量（Égert et al., 2024）。

### （二）学生和成人的出生队列匹配与关联函数

在扩展成人认知技能数据这一研究方向中，受视野局限，目前作者只找到一篇研究：Égert et al. (2024)。因此，本部分内容主要是介绍这一研究的方法思想。

这一方法的思想类似于流量和存量之间的关系。接受教育后的学生作为流量，不断进入劳动力市场，成为成人的一份子；而成人作为存量，在流量的不断融入之下，不断被流量所替代。因此，当今的成人劳动力存量可以看作是由之前时期的学生不断替代后的结果。因此，如果拥有替代全部成人的所有时期流量数据的话（即之前所有历史时期的学生认知技能），就可以利用曾经的流量，估算当今的存量（即当前的成人认知技能）。

如图，假设某个国家$c$的 15 岁学生参与了 1995-2015 年 PISA 测试，该测试得分为流量，该国成人参与了 2017 年的 PIAAC 测试，该测试得分为存量。如果参与了 1995-2015 年 PISA 测试的 15 岁学生在参与测试之后，其认知技能不会再改变的话，我们就可以根据 1995-2015 年 PISA 测试的 15 岁学生在 2017 年时所对应的出生队列，构建出 2017 年时该国的 15-39 岁成人认知技能数据$PISA'2017$。换句话说，图中的$PISA'2017$ 的 15-19 岁的成人认知技能等于参与了 PISA2015 年的 15 岁学生的认知技能，20-24 岁的成人认知技能等于参与了 PISA2015 年的 15 岁学生的认知技能，依次类推（如图中带箭头实线所示）。

然而，事实上，由于中学后教育、在职培训、技能折旧等因素的影响，我们在上面所构建的$PISA'2017$ 并不会真的等于 2017 年 C 国的成人认知技能。为了解决这一问题，需要知道构建的成人认知技能（$PISA'2017$）与真是的认知技能（PIAAC2017）之间的关系。通过构造的$PISA'2017$ 和真实的 PIAAC2017 中各出生队列的对应数据（如图中虚线所示），我们便可以估计出连接转换方程，

行，与 OECD 实施的调查不同，其调查内容主要为阅读能力（Reading Skill）。

基于此方程，便可以实现利用之前时期学生认知技能 PISA 得分来推算当今成人认知技能 PIAAC 的目的。

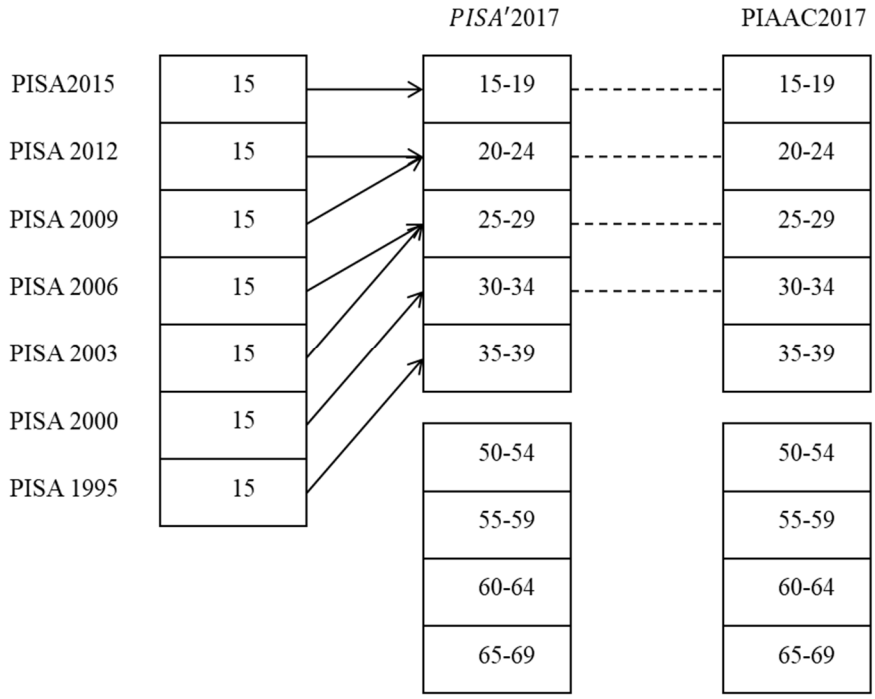



**图 6-1 研究思路简单示意**

根据 Égert et al. (2024)，上述操作过程一般分为三步：

第一步，挑选同时参加 PISA 和 PIAAC 测试的国家，作为锚点国家。按照学生参加测试的时间，推算/虚构出对应时期的成人认知技能数据，如上图中的 *PISA'*2017。

第二步，利用上述国家的虚构成人认知技能数据（如*PISA'*2017）和其真实的成人认知技能（如 PIAAC2017），以出生队列为数据单元，构造转换函数。这是最重要的步骤，转换函数的质量直接关系到之后的转换数据质量。

第三步，应用构建好的转换函数，对参与了很多期 PISA 测试的国家进行数据分析，估计出这些国家各出生队列成人的认知技能得分。再对该国所有出生队列得分进行，便可以获得该国整体成人认知技能的均值或其他统计量。由于估算的成人认知技能数据为 2020 年，因此，参加了 PIAAC 测试的国家也参与了估算。

Égert et al. (2024) 利用开创性地使用历史上每年的人力资本流量，累积计算

获得当前成人的人力资本存量。然而，该方法也有其局限，它更适用于基础教育入学率和升学率较高的高收入经济体，对于经济和教育欠发达的国家和地区来说，因为有大量人口没有接受基础教育，这些国家学生 PISA 得分并不能代表其所对应的出生队列的所有人口，因此采用该方法可能会导致对发展中国家成人认知技能的偏高测量。

通过上述方法，Égert et al. (2024) 构建了 2020 年的 17 个国家 15-64 岁人口与 54 个国家 15-39 岁人口的人力资本存量数据库，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。由于在实践中，该方法的实施严重依赖于样本国家基础教育学生过去是否接受过国际测试及接受了几轮的国际测试，而高收入经济体是国际性学生测试的“常客”，因此 Égert et al. (2024) 构建的样本数据依然是以高收入经济体为主，只包含少量的中等收入经济体，在 15-39 岁人口数据覆盖的国家中，高收入经济体占 74.07%，OECD 国家占 68.52%。

## 七、劳动力市场工资信息：教育质量和人力资本的直接体现

### （一）研究意图：以工资信息中分离出各国教育质量和人力资本信息

理论上，教育质量将直接体现在毕业生的表现上，劳动力在市场上的工资可以作为个人人力资本的直接体现（Lee & Barro, 2001）。因此可以使用劳动力在市场的工资信息来反映各国的教育质量和人力资本。目前有两种方式从工资信息中获得这两类信息：一种是基于明瑟收入方程，用教育收益率来衡量各国教育质量。另一种是基于宏观生产函数，从工资收入中分离出人力资本的影响。

**表 7-1 使用工资信息的两种方法**

| 基础 | 分离出来的测量指标偏向 |
| --- | --- |
| 基于明瑟收入方程 | 教育质量 |
| 基于宏观生产函数 | 人力资本 |

这两个方法的一个问题是各国劳动力的工资表现不仅仅取决于各国的学校教育和人力资本，还取决于外部环境（Lee & Barro, 2001）；并且各国劳动力市场环境的不一致，使得各国之间的工资信息并不可比。劳动力收入的差距并不只反映劳动力人力资本的差异，还有可能反映的是劳动力市场的不同，如全要素生产率、工作信息可获得水平等等。因此，现有研究在样本选择上通常使用某一个国家的劳动力市场上来自其他国家的移民的数据进行研究，这确保了劳动力市场的一致。

在使用移民数据时必须考虑两个问题：首先，移民可能是有选择性的，因

而使得来自某个国家的移民并不能代表这个国家的整体情况。这种选择性，主要来自两个方面，一是移民的自选择，来自低收入国家的受教育程度较高有可能移民至经济发展水平更富裕的国家，因此使用个人能力较高的样本可能会高估这个国家的教育质量；二是移入国家的选择，如美国、英国等国家通常会设定一些条件来筛选移民。其次，技能是否可以完全转移的问题，即在本国学习的技能在移民后是否变得不再适用。如果存在这种情况，所估计出来的市场回报率也并不能代表这个国家真实的教育系统质量。

**（二）处理方法：微观路径下的明瑟收入方程和宏观路径下的人力资本价格**

**1. 明瑟收入方程**

当使用一国移民数据来构建明瑟收入方程时，有如下公式：

$$log\left(W_{c2}^{c1,i}\right) = \gamma_{c2}^{c1} + \mu_{c2}^{c1} * S_{c2}^{c1,i} + \beta * X_{c2}^{c1,i} + \varepsilon_{c2}^{c1,i} \quad （7\text{-}1）$$

其中$c2$指移民移入国家，$c1$指移民来源国家；$log(W)$为个人收入的对数，$S$为个人的受教育年限，$\mu$所关心的估计系数，其代表移民来源国家的教育质量，表示接受一年教育能够带来的收入提升；$X$为其他控制变量。

利用某个国家（如美国）中的移民数据，便可以估计出每个国家的教育质量$\mu^c$。

目前，使用这一方法分离教育质量的研究有限，在移民数据的移民选择和技能转移性两个问题上，现有研究只验证了移民选择和技能转移性问题不会对估计结果产生较大影响（Schoellman，2012）。

**2. 人力资本价格**

这一方法从宏观生产函数出发。标准的生产函数为：

$$Y_c = K_c^{\alpha}(A_c H_c)^{1-\alpha} \quad （7\text{-}2）$$

其中$Y_c$为总产出；$K_c^{\alpha}$为物质资本存量；$A_c$为全要素生产率；$H_c = h_c L_c$为总劳动力投入，$h_c$为人均人力资本，$L_c$为劳动者数量。

对于生产，为了利益最大化，需最大化下式：

$$\max_{H_c} K_c^{\alpha}(A_c H_c)^{1-\alpha} - w_c H_c \quad （7\text{-}3）$$

一阶导数表明单位人力资本的工资（即技能价格）为：

$$w_c = (1-\alpha)z_c，\ where\ z_c = \frac{K_c^{\alpha} {A_c}^{1-\alpha}}{H_c^{\alpha}} = \left(\frac{K_C}{Y_C}\right)^{\frac{\alpha}{1-\alpha}} A_c \quad （7\text{-}4）$$

从中可以看出，各国的技能价格（$w_c$）受各国的全要素生产率以及资本产出比等影响。此外，如果工人以边际产出支付工资，那么工人的收入为：

$$w_{i,c} = w_c * h_{i,c} = (1-\alpha)z_c h_{i,c} \quad （7\text{-}5）$$

由此，

$$\log(w_{i,c}) = \log(z_c) + \log(h_{i,c}) + \log(1-\alpha)$$

$$= \log(z_c) + \log(h_c) + \log(h'_{i,c}) + \log(1-\alpha) \quad （7\text{-}6）$$

其中$\log(h_c) = E_c[\log(h_{i,c})]$。从上述公式中可知，若是知道各国的$z_c$，便可以通过固定效应获得各国的人力资本。

因此，通常这类研究通常分为两步，第一步先估计出$z_c$；第二步利用上式获得各国的人力资本。

下面以最简单的情况来具体说明：当不存在移民选择和技能可转移性问题时，移民的人力资本不随移民发生变化，因此：

$$\log(w_{i,c1}) - \log(w_{i,c2}) = \log(z_{c1}) - \log(z_{c2}) \quad （7\text{-}7）$$

通过估计上式，便可以获得各国的$z_c$，不过这一估计对数据的要求极高，需要有移民前后的工资数据和有多个移民来源国家和移民移入国家[54]。

当存在移民选择和技能可转移性问题时，问题会复杂起来，有研究通过内化移民选择和技能可转移性才解决这一问题（Martellini et al., 2024）。

## 八、人力资本指标分析框架

我们介绍了人力资本指标的演变史，也介绍了演变历程中的五个关键测量指标，然而，现有对指标的分析都比较散乱，没有放入同一个框架之中。

[54] 需要注意的是，当所有的移民都来自一个国家，或者只获得了一个国家的移民数据时，该式只能估计的结果为各国的技能价格与移民国家的差，但这一差值并不影响各国的人力资本估计。

### （一）统一框架中的人力资本指标分析框架

#### 1. 理论框架

根据人力资本理论，人力资本是指存在于人体之中的具有经济价值的知识、技能和体力 (健康状况)等质量因素之和。

$$H = Skills + Knowledge + Health + \varepsilon \quad (8\text{-}1)$$

然而现有的对人力资本的测量，更多地还是对技能的测量，因此，上述公式通常又为：

$$H = Skills + \mu \quad (8\text{-}2)$$

按照通常的划分，技能分为认知技能与非认知技能：

$$H = CognitiveSkills + NoCognitiveSkills + \mu \quad (8\text{-}3)$$

由于测量技术的局限，现有大都集中于认知技能身上，于是有：

$$H = CognitiveSkills + \mu \quad (8\text{-}4)$$

技能，尤其是认知技能的生产可以划分为两个阶段，一是毕业前，二是毕业后。

$$Skills = Skills_{atschool} \ or \ Skills_{afterschool} \quad (8\text{-}5)$$

其中$afterschool$和$atschool$可能存在这巨大的时间间隔。在毕业后，技能主要学生时期的技能（$Skills_{atschool}$）、受到在职教育（$onjob\ education$）、经验（$experience$）和折旧（$aging$）的影响。

$$Skills_{afterschool} = f(Skills_{atschool}, onjob\ education, experience, aging) \quad (8\text{-}6)$$

在毕业前，根据教育生产函数，认知和非认知技能作为产出，其受到学校教育($S$)、家庭投入($F$)、天赋($A$)等因素的影响，根据 Hanushek & Woessmann (2012a)、Hanushek et al.(2015)，公式可以写为：

$$Skills_{atschool} = \lambda F + \phi S + \delta A + \alpha X + \nu \quad (8\text{-}7)$$

毕业前后的划分有两个目的，一是前后所受的影响并不相同；而是为了强调其在时间上存在遥远的距离，防止将同一时间点上的不同阶段的技能相混淆：2025 年的一个成年人的认知技能受其在学生时代所的教育的影响，而非受 2025

年教育的影响。上式子，$S$为学校的影响，其可以看成是学校教育数量和质量的函数：

$$S = f(n, q) \tag{8-8}$$

从另一个角度来讲，学校教育数量和质量也是各类学校投入的产出：

$$n(q) = ExpendPS + Teacher + X + \pi \tag{8-9}$$

其中，$ExpendPS$表示生均经费，$Teacher$表示师资，X 表示其他投入。

**2. 指标角色**

根据以上框架，我们可以将直接-间接、存量-流量、产出-投入、质量-数量、劳动力市场-教育系统放入这个框架之中。

（直接-间接）首先是直接和间接。如果存在所谓人力资本最优测量，它一定是直接指向人的知识和技能，任何不涉及知识和技能、健康等要素的测量指标都只是对人力资本的间接测量。

（流量-存量/劳动力市场-教育系统）其次，我们在谈及一国人力资本时，通常指向的是劳动力市场上的所有成人的人力资本。因此，测量一国人力资本，应当是测量一国人力资本的存量，而还未进入劳动力市场的群体，如教育系统中的学生，其在人力资本存量的来源，是还未发生的流量指标。

（产出-投入）认知技能是技能的重要组成部分。一国某一时期劳动力市场上的成人由不同出生队列的人口组成，根据教育生产函数，**某一时期某一出生队列的人口的认知技能可视为是该年龄段成人在学生时代所获得的认知技能的一个函数；（数量-质量）学生时代所获得的认知技能又可视为是当时该国规定的完成各级各类教育所需最小年限（即教育数量）与其所接受教育质量的一个函数（教育质量可以看作是所接受教育的平均质量）；而教育数量和教育质量又是一国教育长期发展的产物，是该国教育长期投入的结果，是有关各级各类教育投入（如师生比，生均经费）的一个函数。**在不同的情况下，相同的因素（教育数量和教育质量）可以既是投入，也是产出。**如果我们将教育生产视为是一个有关教育投入-产出在不同层次不断转化的动态过程，在最终阶段，一个国家人力资本投资的最终产品应当是成人人口的认知技能，学生时代所获得的认知技能是其投入；在中间阶段，学生时代所获得的认知技能是其产出，学生所接受教育的年限与教育质量是其投入；在最基础阶段，教育质量和教育数量都是产出，而生均经费、入学率、师生比等都是投入。**

### （二）各人力资本测量指标的分类和特点

分析框架下的各人力资本测量指标。在形成框架之前，有必须要先对出现的指标进行总结。

#### 1. 产出-投入、直接-间接、存量-流量、质量-数量

我们根据前面所描述的关键词，按直接-间接、存量-流量、投入-产出和质量-数量对提及的这些指标进行了总结。

**表 8-1 直接-间接和存量-流量划分下的各人力资本测量指标**

| | 存量指标 | 流量指标 |
|---|---|---|
| 直接 | 成人认知技能（均值） | 学生认知技能（均值） |
| 间接 | QAYS（所有人口）<br>从工资信息中分离的人力资本<br>平均受教育年限<br>识字率 | QAYS（某一出生队列）<br>从工资信息中分离的教育质量<br>师生比、生均经费、入学率等 |

**表 8-2 投入-产出、质量-数量划分下的各人力资本测量指标**

| 层次 | 投入 OR 产出 | 数量指标 | 质量指标 |
|---|---|---|---|
| 顶层 | 产出 | 成人认知技能（均值） | |
| | 投入 | 成人在学生时代所获得的认知技能、QAYS（某一出生队列） | |
| 中层 | 产出 | | |
| | 投入 | 从工资信息中分离的教育质量、（年级或年级一致的）学生认知技能（均值） | 平均受教育年限 |
| 底层 | 产出 | | |
| | 投入 | 师生比，生均经费 | 入学率 |

#### 2. 劳动力市场和教育系统

虽然以上指标都可以用于测量一国人力资本，但在其偏向上，一些指标更倾向于测量一国教育系统指标。在文中用词上，我们尽量区分了人力资本和教育质量，在下表中，总结了这些指标的倾向。对此进行区分的重要和必要原因出自对时间维度信息的关注，教育系统对整体人力资本的影响是渐进、缓慢而滞后的，因此，如果要进行对应，教育系统指标的时间要早于人力资本测量指标，而现有研究的一些处理忽略了这一点。

我们认为，平均受教育年限既是人力资本测量指标，也是教育系统指标，因为这里是纯粹使用教育来衡量人力资本；学生认知技能是教育系统指标，因为其测量的对象大多是初级教育和中等教育的学生；调整后的平均受教育年限，

依据平均教育年限是整体人口还是刚进入劳动力市场的出生队列人口，分别偏向人力资本和教育系统，但需要知道的是，从理论上讲，用于调整的教育质量是从学生认知技能数据中获得的，因为这一方法理论应当是偏向教育系统；成人认知技能由于是劳动市场上人口的认知技能，正如本文所提倡的，其是人力资本最优的代理指标；劳动力市场上的工资信息，是用来分离出教育质量，因此其也偏向于教育系统。


**图 8-1 各测量指标在人力资本-教育系统划分下倾向示意图**

### （三）基于统一框架下的各人力资本测量指标

#### 1. 人力资本测量指标优劣评价原则

前文的人力资本指标演变史本就反应了对人力资本指标评价的倾向，本文将这些倾向总结为五个评价原则：“全面测量优于单维测量”“直接测量优于间接测量”、“存量测量优于流量测量”、“产出测量优于投入测量”和“质量测量优于数量测量”。

第一，人力资本内涵丰富，即包括技能、也包括知识、健康等，现有对人力资本的测量大多集中在技能，尤其是认知技能上，技能或认知技能只是人力资本的一个维度，而非全部，最优的测量应当包含所有的人力资本内涵。

第二，如果存在所谓人力资本最优测量，它一定是直接指向人的知识和技能，任何不涉及知识和技能的测量指标都只是对人力资本的间接测量，也就有“直接指标优于间接指标”。从这一角度看，如果我们要对一国教育人力资本进行测量，只要对该国成人认知技能进行测量即可；如果我们要对一国教育系统质量进行调查，也只要对该国学龄儿童在接受正式教育前后的认知技能增值变化进行测量即可。

第三，人力资本能够促进经济增长，虽然没有明确指明，但这里的人力资本应当所有人的人力资本，因此在理论上，应当直接对劳动力市场上的所有成

人进行测量（即存量测量），而非对学生测量（即流量测量）。一国教育发展对人口人力资本的影响是滞后的。已有研究也证实了世界上大部分国家的基础教育学生认知技能与成人认知技能在分布上通常会发生较大改变，因此用当期学生认知技能来代理同期成人认知技能，这一做法是值得商榷的（黄斌 等，2024）。于是“存量指标优于流量指标”。

第四，在将**教育投入-产出在不同层次不断转化的动态过程**中，虽然各种投入和产出指标都曾经被用来衡量一国教育人力资本。但我们应当知道，对产出的测量理应高于对投入的测量，因为将所有的投入都看作是有效的生产性投资是不合适的（Hanushek，2003），投入能够产生多大的产出也是存在不确定性的（Schoellman, 2012; Hanushek, 2003）。在这一情况下，直接对产出进行测量是更正确的做法，因此“产出指标优于投入指标”。

第五，随着人力资本理论被广为传播和接受，全球，几乎每一个国家发展教育、发展人力资本。发展教育应当同时关注教育质量和教育数量，提升教育质量和教育数量都是为了提升最终的认知技能产出。如果一个国家为民众提供极为低质的教育，哪怕该国国民平均教育年限达 16 年以上，该国国民认知技能水平也不会太高，这样的教育投资是低效的，也不会对外部社会经济发展产生的经济价值。同样，如果一个国家为民众提供数量有限的教育，哪怕该国的教育质量再优质，其国民认知技能水平也不会太高，因为一年教育所能产生的影响不可能太大，这样的教育投资对外部经济社会产生的经济价值也终将是有限的。因此，对于国家长期发展来说，教育数量扩招和质量提升是同等重要的。然而，随着投资人力资本这一普遍观念在全球范围内传播，很多信息发生了改变和扭曲，丧失了很多本质内容和力量（Hanushek & Woessmann, 2015）。政策制定者和学者并没有真正重视人力资本的本质：知识、技能等，而是把注意力放在了与教育程度相关的代理指标上，如平均受教育年限、入学率等数量维度的指标。以平均受教育年限、入学率等数量指标作为政府政策着力点的现实事实告诉我们，一些国家（如拉丁美洲国家）在自己和其他国家的帮助下扩大了教育机会、提高了平均受教育年限，但经济增长依然缓慢，没有明显赶上发达国家的痕迹（Hanushek & Woessmann, 2008）。这些指标掩盖了各国教育质量的差异；并且“Schooling is not learning”（Pritchett, 2013; World Bank, 2018; Kaffenberger & Pritchett, 2017; Filmer et al., 2020; Angrist et al., 2021），只有将学校教育（Schooling）转化为学生的有效学习（Learning）与真实技能（Skill）的提升，才能使得教育发展具有生产性，进而对一国长期经济增长产生持续的内生推动作用（黄斌 和 云如先, 2023）。因此，在这样的背景下，“质量指标优于

数量指标”。

**2. 各人力资本指标分析**

从直接与间接测量的角度看，（学生/成人）认知技能指标优于 QAYS 或 LAYS、受教育年限等指标；从存量和流量角度看，成人认知技能指标、受教育年限优于学生认知技能、入学率指标；从产出与投入角度看，成人认知技能指标等指标优于师生比、生均经费等指标；从质量和数量角度看，学生认知技能指标又优于受教育年限、入学率等指标。综合以上对比，成人认知技能应是衡量国家教育人力资本的最佳指标，它与其他指标相比拥有着绝对的优势。

其次，我们认为，如果目标是测量人力资本的话，优先劳动力市场指标，有直接测量的话优先使用直接测量，若无直接测量的话，优先存量测量，若无存量测量，也应优先产出指标，在投入指标中，优先质量测量。

（1） 成人认知技能、受教育年限和学生认知技能三者关系的再讨论

需要再次说明关于成人认知技能、平均受教育年限和学生认知技能三者的关系的认识。这三者是现有研究中的焦点，增加对其的认知有助于了解现有研究。

根据分析框架，可有以下公式：

$$\text{成人认知技能} = f(\text{对应学生时代认知技能})$$

$$= g(\text{教育质量} * \text{教育数量}(\text{受教育年限}))$$

$$\text{学生认知技能} == g(\text{教育质量} * \text{教育数量}(\text{年级对应的受教育年限}))$$

成人认知技能是成人对应学生时代的认知技能的函数。需要注意的是，对应学生时代的认知技能与学生认知技能有巨大的差别，首先，两者在时间上并不一致，当今的成人所对应的学生时代所处的时间要在很多年前，而当今的学生认知技能是当今的；其次，在概念上，对应学生时代认知技能是其接受教育后的最终结果，这是理想的，事实上并无测量，如果测量的话，应该在其最终离开学校时进行测量，由于不同人离开学校的时间不同，这使得这件事难以实行；而实际测量的学生认知技能，由于测量原因，更多地是指某一年龄或年级的在校生，其依然在接受教育，其认知技能并非是最终的结果。

由于以上区分，因此，严格而言，下式并不成立：

$$\text{成人认知技能} = g(\text{教育质量（学生认知技能）} * \text{教育数量}(\text{受教育年限}))$$

其次，对于学生认知技能而言，其可以作为教育质量，只要其教育数量是一致的，也就是说，是对同一年龄或年级的学生进行的评估，否则的话，学生认知技能并不等于教育质量。

## 九、总结、评价与展望

可靠而精确地测量各国人力资本，是开展宏观教育政策研究的基础性工作。本文沿着人力资本测量的“演变史”，系统介绍受教育年限、学生认知技能、QAYS、成人认知技能和劳动力市场工资信息中分离出的教育质量和人力资本等人力资本测量指标，包括它们的存在问题、构建原理和方法与常用国际数据库等。在系统介绍完国家教育人力资本的主流指标后，有必要对这四种指标进行更深度的总结和对比研究，分析它们各自存在的问题，凝练评判各类测量指标优劣等次的基本原则，并指出未来研究的发展方向。

### （一）总结：人力资本测量指标的发展规律

从人力资本测量指标发展的过程来看，可以明显发现以下几点规律，一是优先以产出指标，而非投入指标来衡量教育获得或人力资本，平均受教育年限和认知技能都是直接的产出端指标。从产出端而非投入端衡量教育系统或人力资本是结果为导向的结果，是对“投入能否带来产出？能够带来多大产出？”的思考结果。

二是从重视数量指标转向越来越重视质量指标，人力资本衡量指标从平均受教育年限转移至认知技能充分地体现了这一点。这既是受现有以平均受教育年限等数量指标为目的的政策没有带来预期的经济增长的影响，也是受各教育系统之间教育质量的差异和人们对学校教育并不等于学习（Schooling is not same as learing）认识的影响。

三是在宏观研究中，研究者从只考虑数量，到将教育数量和教育质量同时加以考虑。但需要注意的是，对于个体而言，这意味既要接受足够多数量的教育，也要求所接受的教育质量足够好；对于国家而言，这意味着“大量的受教育人口，但教育质量不足”和“卓越的教育质量，但有限的受教育人口”都会限制一国生产力的发展。

四是，虽然近年来在认知技能数据上，对流量指标（学生认知技能）关注过多，但有回归存量指标（成人认知技能）的趋势。人力资本存量一般指所有

劳动力的人力资本，相对人力资本流量而言，其与经济等相关指标的因果关联更高（De La Fuente & Doménech，2024）。

### （一）评价：人力资本测量指标和数据存在局限

在第八章的分析之中，我们假定各人力资本指标都是完美的，其测量不存在问题。然而，从第三章至第第七章也可以知道，由于数据构建，

1. 受教育年限

为了获得各国之间可比的平均受教育年限，在数据构建时需要进行统一和估计，比如将各国不同的教育阶段划分统一处理为几个广泛的阶段。由于各个国家公布的统计数据是基于各自的通国际规范标准的，所以以上操作会使得构建的数据结果和各国公布的统计数据存在出入，因而不宜直接将构建的数据和各个国家公布的数据直接进行比较。此外，最大的缺陷也正如其被批评的，其无法反映所接受教育的质量的不同。

2. 认知技能

用认知技能来衡量教育质量或者人力资本已成为当前的重要研究和发展方向，然而不可否认的是，就目前数据而言，还存在不少局限。

就测量内容而言，目前还没有测试项目可以覆盖和测量决定一国创新能力和劳动力生产率的所有能力和技能，包括非认知技能、大学和工作场所获得的技能、科学家和高水平技术人才的高度专业化和复杂化的知识和技能等（De La Fuente & Doménech，2024）。就测量对象而言，现有的国际性认知技能测试项目更多的是测量学生认知技能信息，因此只能够用来衡量初等教育或中等教育的质量，目前尚无衡量高等教育阶段质量的认知技能测试项目数据；对于拥有较低入学率的国家而言，这些学生认知技能测试项目得分可能只代表参与测试项目的出生队列人群中的部分人的知识和技能，而非整个出生队列的知识和技能；另外，学生还未进入劳动力市场，因此用来替代劳动力作为整个国家的教育质量或者人力资本测量指标还尚有不足，在理论上就存在严重的内生性和互为因果的问题（黄斌 等，2024；De La Fuente & Doménech, 2024）[55]。就测量国家和时间而言，我们也还远没有足够的信息来衡量全球大多数国家长时间的人力资本（和变动）。即使是学生测试项目，很多国家也只参与了最近几次的国际

[55] 在这种情况下，使用劳动力人口的平均受教育年限的研究可能受反向因果问题的影响相对更小（De La Fuente & Doménech，2024）。当然，最直接的办法还是直接使用成人认知技能（黄斌 等，2024）。

性测试项目，因而只有少数国家具备长期序列的数据。另外，现有的认知技能数据的质量还有待进一步提升，目前虽然研究都提及了内容相同假设不严格满足情况下可能存在的风险，但尚无研究考虑内容相同假设不严格满足时，如何通过一定的方法解决测试内容差异问题，获得内涵更加一致的转换数据，以提高转换数据的质量。

在学生认知技能作为教育获得这一用途上，学者通常只考虑认知技能得分的可比，而忽略学生的一致（如年龄）。

3. 改造平均受教育年限

改造平均受教育年限将数量和质量维度包含在同一个指标内，这一方面既没有抛弃常用的平均受教育年限，也加入了认知技能这一质量维度指标。

然而，这一指标包含着较大的构建误差。研究的实施都依赖对认知技能生产函数的假定（不可否认地是，这一研究加深了人们对认知技能生产函数的了解），而对认知技能生产的了解不足将限制这一方法所得最终结果的精确性；克服对认知技能生产函数的了解不足，则需要测量每个年龄群体的认知技能，并观察普遍的认知技能增长规律，这其中既要考虑教育的影响，也应当考虑到生理和心理的成熟，对于这些，研究者还需要付出更多的努力。另外，当使用成人认知技能来估计教育质量时，由于认知技能存在折旧，以及在职培训、干中学等因素的影响（De La Fuente & Doménech, 2024），这使得获得干净的教育系统质量更加困难，在估计时误差更大。此外，容易被忽视的是认知技能在测量上也同样存在测量误差，这一误差会累积至改造后的平均受教育年限上。

其次，该指标的作用有限。在各教育阶段教育质量一致时，若想衡量教育质量，某一年级的学生认知技能得分是较好的衡量指标；在各教育阶段教育质量不一致时，若想衡量教育质量，学生认知技能得分的增值变化是较好的衡量指标；当衡量教育数量时，平均受教育年限时较好的指标；当衡量最终教育产出时，最终的成人认知技能得分是最好的指标。在这一框架下，无疑并不需要改造平均受教育年限，为了更好的服务政策目标，应当加强教育统计和调查，准确把握人口受教育年限数据以及各年龄段人口的认知技能数据。

4. 市场上的工资信息

虽然从劳动力市场上的工资信息中分离出教育质量或人力资本在拥有移民数据的情况下切实可行，但这一方法因其自身的特殊性使得其在与平均受教育年限和认知技能数据比较时处于劣势地位，致使其在衡量教育质量或人力资本

这一点上的应用有限。

首先，在考虑因素上，使用劳动力市场上的工资信息使得研究还需要考虑学生进入劳动力市场后的一些因素，这使得在理论上这类研究就比之前的研究更难实施。

其次，在实践中的数据使用上，其对数据的要求更高。一方面这一研究需要使用移民数据，在数据获取上本身难度就较大；并且其结果易受不同来源数据的影响，虽然不同国家的移民数据得到的结果有很大的相关性，但却无法否定其结果的不稳定。另一方面其衡量的教育系统质量是进入劳动力市场的个体的加权值，由于这类研究都假设教育系统质量不变，以此忽视对时间的考虑，因此，理想状态下，应该包括足够的各时期的具有代表性的样本。如果所观测某些国家的个体样本是早进入劳动力市场的，其市场回报率衡量的可能是早先该国教育系统的质量，与其他国家实际上可能是不可比的。

最后，在时间变动上，这类研究得到的指标无法衡量教育质量和人力资本在时间上的变动。这类研究都是基于不同毕业时间和相近工作时间的移民收入数据，得到一个不随时间变化的国家教育质量或人力资本的估计量。然而，一个国家不同时期的教育质量是不一样的，使得接受相同教育数量的个体回报率、以及不同时期接受教育获得的人力资本都不尽相同，同时这种方法很难比较教育系统质量或人力资本在时间上的差别。

由于上述限制，使得该类研究的教育质量或人力资本难以用于政策实践中。

### （二）展望：人力资本测量指标使用的建议

第一，分布总会比单一的均值包含更多的信息。这一点不管是在教育成就或是在认知技能上都是一样的。在教育成就上，受教育程度的分布（达到某一教育阶段的人口的比重）要比单一的平均受教育年限信息更多；在认知技能上，认知技能分布（均值、分位点值、标准差、偏态）要比单一的认知技能均值要信息更多。对分布的重视，虽然在 Hanushek & Woessmann (2012a)、黄斌和云如先（2023）、黄斌 等（2024）的推动下有所发展，但还有待深入探索。使用分布信息既是对现有信息的充分利用，更是在实证中人力资本理论的创新发展，意味着研究从简单的人力资本总体水平向人力资本差异特征、结构特征等更丰富内涵的延伸。此外，在学生认知技能可比数据库构建中，各种方法应用于其他分布特征转换的适用性和精确性还需要进一步分析和讨论。

第二，不应忽视教育质量随时间的变化。在教育快速扩张，尤其是高等教

育快速扩张的背景下，研究人力资本随着时间的变化尤为重要。教育质量保持不变是大多数研究的基本假设（De La Fuente & Doménech，2024），然而教育快速扩张时期，教育的数量和质量都会发生巨大变化，这时候在人力资本的测量上不考虑时间维度，将影响研究结果的可靠性；简单的实证分析也表明，在很多国家，教育质量并没有长期保持稳定（Hanushek & zhang，2009）。若要实现对一国教育人力资本存量的时间变化，就不能只关注各国单个历史时期的人口认知技能均值水平，必须要对不同历史时期的不同出生队列人口认知技能进行分组统计描述，并且充分了解各国教育机制体制重大改革的历史沿革，就这些改革对不同历史时期国家教育人力资本存量变化的因果效应进行估计。

第三，成人认知技能虽然是一国人力资本最优的测量指标，但用认知技能来衡量教育质量或人力资本，还需要进一步加强对其的认识。不论是学生认知技能，抑或是成人认知技能，都不仅仅是学校教育的产物，校外教育的影响对认知技能的影响究竟贡献了多少，至今尚无研究能够给出准确回答。此外，虽然研究常使用学生认知技能作为各国教育质量的衡量指标，但学生认知技能的生产也应是教育数量和教育质量两种投入的共同产出，因此，应当使用相同年级（此时接受教育的年限相同）的学生的认知技能，而不应当多个年级的数据混合使用。虽然认识到这一点并不太影响之前的研究，但对此的强调却并不多见。

## Supplementary data

在附录中，我们给出了各方法的文献汇总，常用数据库结果以及部分文献的方法详细介绍。

### 参考文献


[1] 黄斌, 云如先. (2023). 教育发展何以强国——基于 1960—2020 年认知技能国际可比数据的实证分析. 教育研究, 44(10), 125-136.

[2] 黄斌, 云如先, 吴凯霖. (2024). 认知技能分布对国民经济增长的影响：教育强国的新证据. 华东师范大学学报(教育科学版), 42(9), 13-32.

[3] Altinok, N., Angrist, N., & Patrinos, H. A. (2018). Global Data Set on Education Quality (1965–2015). *Policy Research working paper*, Washington, D.C.: World Bank Group.

[4] Altinok, N., & Diebolt, C. (2024). Cliometrics of Learning-Adjusted Years of Schooling: Evidence from a New Dataset. *Cliometrica,* 18(3), 691-764.

[5] Altinok, N., Diebolt, C., & Demeulemeester, J. L. (2014). A New International Database on Education Quality: 1965–2010. *Applied Economics*, 46(11), 1212-1247.

[6] Altinok, N., & Murseli, H. (2007). International Database on Human Capital Quality. *Economics Letters*, 96(2), 237-244.

[7] Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring Human Capital Using Global Learning Data. *Nature*, 592(7854), 403–408.

[8] Angrist, N., Evans, D., Filmer, D. P., Glennerster, R., Rogers, F. H., & Sabarwal, S. (2020). How to Improve Education Outcomes Most Efficiently? A Comparison of 150 Interventions Using the New Learning-Adjusted Years of Schooling Metric. CDG *Working Paper*.

[9] Balaj, M., Henson, C. A., Aronsson, A., et al. (2024). Effects of Education on Adult Mortality: A Global Systematic Review and Meta-Analysis. The Lancet Public Health, 9(3), e155-e165.

[10] Barro, R. J., & Lee, J. (1993). International Comparisons of Educational Attainment, *Journal of Monetary Economics*, 32(3), 363-394.

[11] Barro, R. J., & Lee, J. (2001). International Data on Educational Attainment: Updates and Implications. *Oxford Economic Papers*, 53(3), 541-563.

[12] Barro, R. J., & Lee, J. (2013). A New Data Set of Educational Attainment in the World, 1950–2010. *Journal of Development Economics*, 104(September), 184-198.

[13] Barro, R. J., & Lee, J. (2015). Education Matters: Global Schooling Gains from the 19th to the 21st Century. Oxford, UK: Oxford University Press.

[14] Bauer, R., Potančoková, M., Goujon, A., & K.C., S. (2012). Populations for 171 Countries by Age, Sex, and Level of Education around 2010: Harmonized Estimates of the Baseline Data for the Wittgenstein Centre Projections. IIASA Interim Report IR-12-016.

[15] Braun, H. I., & Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland & D. B.

Rubin (Eds.), *Test equating* (pp. 9-49). New York: Academic.
[16] Cohen, D., & Soto, M. (2007). Growth and Human Capital: Good Data, Good Results. *Journal of Economic Growth*, 12(1), 51-76.
[17] Das, J. & Zajonc, T. (2010). India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement. *Journal of Development Economics*, 92(2), 175-187.
[18] De La Fuente, Á., & Doménech, R. (2000). Human Capital in Growth Regressions: How much Difference Does Data Quality Make? *Working Paper*.
[19] De La Fuente, Á., & Doménech, R. (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make? *Journal of the European Economic Association*, 4(1), 1-36.
[20] De La Fuente, Á., & Doménech, R. (2015). Educational Attainment in the OECD, 1960–2010. Updated Series and a Comparison with Other Sources. *Economics of Education Review*, 48(October), 56-74.
[21] De La Fuente, Á., & Doménech, R. (2024). Cross-Country Data on Skills and the Quality of Schooling: A Selective Survey. *Journal of Economic Surveys*, 38(1), 3-26.
[22] Égert, B., De La Maisonneuve, C. & Turner, D. (2024). A New Macroeconomic Measure of Human Capital Exploiting Pisa and Piaac: Linking Education Policies to Productivity. *Education Economics*, 0(0), 1-17.
[23] Filmer, D., Rogers, H., Angrist, N., & Sabarwal, S. (2020). Learning-Adjusted Years of Schooling (LAYS): Defining a New Macro Measure of Education. *Economics of Education Review*, 77(August), 101971.
[24] Gethin. (2023). Distributional Growth Accounting: Education and the Reduction of Global Poverty, 1980-2022. Job Market Paper.
[25] Glawe, L., & Wagner, H. (2022). Is Schooling the Same as Learning? – The Impact of the Learning-Adjusted Years of Schooling on Growth in a Dynamic Panel Data Framework. *World Development, 151(2022)*, 105773.
[26] Goujon, A., K.C., S., Speringer, M., et al. (2016). A Harmonized Dataset on Global Educational Attainment Between 1970 and 2060 – an Analytical Window into Recent Trends and Future Prospects in Human Capital Development. *Journal of Demographic Economics*, 82(3), 315-363.
[27] Gust, S., Hanushek, E. A., & Woessmann, L. (2024). Global Universal Basic Skills: Current Deficits and Implications for World Development. *Journal of Development Economics*, 166(January), 103205.
[28] Hastedt, D. & Desa, D., (2015) "Linking Errors Between Two Populations and Tests: A Case Study in International Surveys in Education". *Practical Assessment, Research, and Evaluation* 20(1), 14.
[29] Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal*, 113(485), 64-98.
[30] Hanushek, E. A., & Kimko, D. D. (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review,* 90(5), 1184-1208.
[31] Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3), 607-668.
[32] Hanushek, E. A., & Woessmann, L. (2012a). Do Better Schools Lead to More

Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth*, 17(4), 267-321.
[33] Hanushek, E. A., & Woessmann, L. (2012b). Schooling, Educational Achievement, and the Latin American Growth Puzzle. *Journal of Development Economics*, 99(2), 497-512.
[34] Hanushek, E. A., & Woessmann, L. (2015). Universal Basic Skills: What Countries Stand to Gain. Organisation for Economic Co-operation and Development, Paris.
[35] Hanushek, E. A., & Zhang, L. (2009). Quality‐Consistent Estimates of International Schooling and Skill Gradients. *Journal of Human Capital*, 3(2), 107-143.
[36] Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics*, 25(2), 133-183.
[37] Kaarsen, N. (2014). Cross-Country Differences in the Quality of Schooling. *Journal of Development Economics*, 107(March), 215-224.
[38] Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics,* 9(1), 25-44.
[39] Kolen, M. J., & Brennan, R. L. (2014). Test Equating, Scaling, and Linking: Methods and Practices. New York: Springer.
[40] Krueger, A. B., & Lindahl, M. (2001). Education for Growth: Why and for Whom? *Journal of Economic Literature*, 39 (4), 1101-1136.
[41] Lee, J., & Barro, R. J. (2001). Schooling Quality in a Cross–Section of Countries. *Economica*, 68(272), 465-488.
[42] Lim, S. S., Updike, R. L., Kaldjian, A. S., et al. (2018). Measuring Human Capital: A Systematic Analysis of 195 Countries and Territories, 1990–2016. *The Lancet*, 392(10154), 1217-1234.
[43] Lucas, R. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22(1), 3-42.
[44] Lutz, W., Goujon, A., K.C., S., & Sanderson, W. C. (2007). Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000. *Vienna Yearbook of Population Research*, 193-235.
[45] Martellini, L., & Schoellman, T. (2012). Human Capital and Development Accounting: New Evidence from Wage Gains at Migration. *The Quarterly Journal of Economics*, 133(2), 665-700.
[46] Patel, D., & Sandefur, J. (2020). A Rosetta Stone for Human Capital. *CGD Working Paper*.
[47] Patrinos, H. A., & Angrist, N. (2018). Global Dataset on Education Quality: A Review and Update (2000-2017). *Policy Research working paper*, Washington, D.C.: World Bank Group.
[48] Pritchett, L. (2013). The Rebirth of Education: Schooling ain't Learning. Washington, D.C.: CGD Books.
[49] Reardon, S. F., Kalogrides, D., & Ho, A. D. (2021). Validation Methods for Aggregate-Level Test Scale Linking: A Case Study Mapping School District Test Score Distributions to a Common Scale. *Journal of Educational and Behavioral Statistics*, 46(2), 138-167.

[50] Reiter, C., Özdemir, C., Yildiz, D., Goujon, A., Guimaraes, R., & Lutz, W. (2020). The Demography of Skills-Adjusted Human Capital. *Working Paper*.
[51] Romer, P. (1986). Increasing Returns and Long-run Growth. *Journal of Political Economy*, 96(5), 1002-1037.
[52] Sandefur, J. (2018). Internationally Comparable Mathematics Dcores for Fourteen African Countries. *Economics of Education Review*, 62(February), 267-286.
[53] Schoellman, T. (2012). Education Quality and Development Accounting. *The Review of Economic Studies*, 79(1), 388–417.
[54] Schultz, T. M. (1961). Investment in Human Capital. *American Economic Review*, 51(1), 1-17.
[55] Singh, A. (2014). Emergence and Evolution of Learning Gaps across Countries: Linked Panel Evidence from Ethiopia, India, Peru and Vietnam. *CSAE working paper*.
[56] Speringer, M., Goujon, A., K.C., S., Potančoková, M., Reiter, C., Jurasszovich, S., & Eder, J. (2019). Global Reconstruction of Educational Attainment, 1950 to 2015: Methodology and Assessment. *VID Working Paper*.