How to Measure Human Capital of A Country?

From Years of Schooling to Cognitive Skills and Wage Information from Market
(Online version)

Version: 2024-02-25

Ruxian Yun

(Institute of Education, Nanjing University 210023 ）

Description :

This is the online English version of the paper "How to measure a Country's human capital".

Translated by a translation application, but no proofreading has been done yet .

Compared with the journal version, the online version mainly differs in the following aspects: firstly, related concepts, such as educational human capital in the online version and human capital in the online version. Secondly, the online version contains content that has been reduced due to layout limitations and ease of understanding. The third part is the summary and commentary section, which is the first author's rough idea. materials). Readers who find it difficult to read the original literature or want to save energy can read the appendix.

In addition, due to limitations in abilities and perspectives, if readers find any omissions, errors, or inappropriate aspects in the article, they are welcome to correct them. The author will humbly accept and promptly correct and update this document. Readers are also welcome to provide their own ideas, supplement the article, and update the content of the article. I will express my gratitude to those who have contributed to this article in specific sections.

As it is an online version, the author has slightly relaxed their wording and is not academic in nature. Please be patient.

Disclaimer: (1) This article is the author's knowledge sharing and intended for public welfare purposes. (2) Reproduction (other official account, websites, etc.) is prohibited. (3) This article should not involve any rights issues. If you have any questions, please contact the author of this article.

Author's contact email: jssyyrx@163.com .

This article can currently be obtained from the following website: https://yrx-china.github.io/publications/2024-11-18-HMACHC

Thank you :

Thank you to the external review experts of the Peking University Doctoral Forum on Educational Economics for their valuable opinions. Thank you to He Peiyun , Liu Cheng, Liao Bin, Wang Ling, and Wu Kailin from the Education Research Institute of Nanjing University for their contributions during the revision process of this article.

How to Measure Human Capital of A Country? 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。

I. Introduction

Human capital is one of the most important factors in promoting national economic growth. Therefore, to complete a macroeconomic econometric study or other research focusing on economic and social development [2], it is necessary to first accurately measure a country's human capital. Although human capital investment can be achieved through education, migration, health and other channels, education is the most important and main channel for a country to accumulate human capital. Therefore, in actual research, human capital is often narrowly defined as simply coming from education investment. Therefore, previous studies often only use education-related indicators to measure [3]human capital (Reiter et al., 2020) [4]. To be precise, the results obtained from this operation point more to educational human capital rather than complete human capital.

The report of the 20th National Congress proposed the "education, science and technology, and talent" three-in-one development strategy, which fully reflects the fundamental and strategic role of education in promoting the construction of a strong country. The strategy of building a strong country through education emphasizes the productive function of education in external social and economic development, and requires education to improve its support and contribution to the realization of common prosperity and China's modernization strategy. The development of education has always been considered the most important investment method for a country to acquire and accumulate human capital, and human capital is an important source of long-term economic growth (Lucas, 1988; Romer, 1986; Schultz, 1961). Therefore, a most basic and important question naturally arises : How should a country's human capital be measured?

The measurement of human capital is the foundation of macroeconomic policy and quantitative research. If we cannot solve the measurement problem of human capital and ensure its measurement accuracy and validity, there is no way to talk about the validity of subsequent research. Existing studies have pointed out that the failure

[2] Mainly focuses on the research of development accounting, growth accounting and empirical macro growth equations.

[3] This article does not specifically distinguish between the usage of the words measure, measure, and metric.

[4] Measuring education systems is often considered to be measuring human capital, which has led to some confusion in measurement and related terminology, such as measuring the quality of education systems is considered to be measuring the quality of human capital.

of previous macroeconomic research to obtain satisfactory conclusions is closely related to the large measurement bias of human capital used (Krueger & Lindahl , 2001; De La Fuente & Doménech , 2006). On the other hand, the academic community has always adopted a "take-it-as-it-is" approach to human capital measurement indicator data between countries , knowing very little about the details of the construction of the human capital measurement indicators used . As a result, they often mistakenly "force" analysis on the originally incomparable educational human capital data of different countries, and the international comparison results obtained are often not credible.

In addition, in recent years, especially after 2010, human capital measurement indicators have also made great progress. In addition to the traditional years of education, indicators such as student cognitive skills, quality-adjusted years of education , and adult cognitive skills have emerged and have been applied in some studies, and conclusions with important policy implications have been obtained. The above situation makes it urgent and important to systematically review and compare the construction methods of existing human capital measurement indicators . However, to the best of the author's knowledge, there is currently no research in China focusing on this issue, and the sporadic foreign literature on this issue is not comprehensive (De La Fuente & Doménech , 20 24).

In view of this, this paper first sorts out the evolution of national education human capital measurement indicators (Chapter 2), and then systematically introduces the commonly used human capital measurement indicators along the development context (Chapters 3 to 7), including the data problems of various indicators (what problems?), data construction principles and methods (how to solve them?). Finally, it summarizes the general understanding of these indicators and the overall development law, and constructs a unified analysis framework based on human capital theory and education production function theory, placing all the indicators mentioned in this paper under this framework (Chapter 8). Finally, it summarizes the development law of human capital measurement indicators, evaluates their limitations, and discusses the future development direction of national human capital research (Chapter 9). It is hoped that with the help of this article, researchers can strengthen their understanding of education human capital measurement indicators and data, and help readers better apply these relevant indicators and data to research practice [5].

II. The Evolution of National Human Capital Measurement Indicators

Since the 1960s, UNESCO has systematically collected data such as enrollment rates at all levels of education in various countries. Therefore, early macroeconomic econometric studies often used enrollment rates to proxy for human capital in various countries and used them to empirically analyze the impact of human capital on total productivity (Hanushek & Kimko , 2000 ; Barro & Lee, 2013) . The advantage of [6]the enrollment rate indicator is that it is easy to obtain, but it is obvious that it deviates from the connotation of human capital theory: human capital is a stock concept, while the enrollment rate is an educational flow indicator. The response of human capital to the enrollment rate is gradual and has a very large lag. Therefore, the enrollment rate of a country in a certain period cannot reflect the accumulated level of human capital stock of the country's adult labor force at that time. In comparison, from a theoretical perspective, years of schooling is an indicator of educational output, so a country's mean years of schooling (MYS) can better reflect the country's human capital than input indicators such as enrollment rate. In addition, measuring the years of schooling of a country's adult labor force and calculating the country's mean years of schooling is actually measuring the country's human capital stock. Driven by this concept , international organizations and a large number of researchers have begun to try to build a database of the average years of schooling of the labor force in various countries to achieve a more direct measurement of the human capital stock of various countries. The average years of schooling has gradually replaced the enrollment rate and become a common indicator for measuring a country's human capital (De La Fuente & Doménech , 2006).

The average years of education as a measure of a country's human capital also has many defects. With the widespread use of this indicator, two criticisms have emerged. The first criticism emphasizes that the average years of education data provided by different sources have large measurement errors, which makes the empirical analysis of economic growth using the average years of education as a measure of a country's

[5] This article also provides a literature summary of various methods and common database results, see the appendix for details.

[6] In the early days, literacy rates were also widely used in empirical research, but they are more difficult to obtain. For a detailed discussion of literacy rates and school enrollment rates, see Berro & Lee (1993) .

human capital stock often obtain disappointing results (Krueger & Lindahl , 2001 ; De La Fuente & Doménech , 2006). The second criticism points out that the average years of education only measures the amount of education received by a country's citizens, while ignoring the huge differences in the quality of education among countries. Macroeconomic econometric research often uses countries and regions as samples for analysis, and the quality of education in different countries and regions obviously varies greatly, which means that the human capital generated by receiving the same number of years of education in different countries should be different. Measuring human capital by years of education implicitly assumes that receiving the same number of years of education in different countries and regions can obtain the same educational results, which is obviously inconsistent with the facts. For example, few people actually think that one year of high school education in the United States is equivalent to one in Egypt (Hanushek & Kimko , 2000). This means that even if the average years of education in two countries are the same, there is a huge difference in the quality of education received by the populations of the two countries, so there is also a huge difference in the stock of human capital in the two countries (in fact, no one dares to say that the quality of education in the two countries is the same at any time). Based on this view, the average years of education as a quantitative indicator of a country's education obviously cannot well reflect a country's human capital.

Although the first criticism has gradually decreased with the improvement of data, the second criticism has become increasingly strong. Researchers have realized that the ideal human capital measurement indicator should reflect both the quantity of a country's human capital and the quality of a country's human capital. In order to control the differences in education quality among countries in research, many studies use indicators such as teacher-student ratio and per capita funding to represent the education quality of various countries. However, these indicators are all factors of educational production input. It is inappropriate to regard all these inputs as effective productive investment (Hanushek , 2003), and there is uncertainty about how much output the input can produce (Schoellman, 2012; Hanushek, 2003) . There is still a huge controversy in the academic community today about whether "education investment is useful or useless." For example, compared with urban schools, rural schools have a smaller student-teacher ratio, but the quality of education in rural schools is usually worse than that in urban schools; similarly, due to the scale effect, the per capita cost (funding) of small-scale schools is relatively high, but the quality of education in small-scale schools is not necessarily higher than that of other schools.

Relatively speaking, measuring a country's education quality and human capital from the perspective of educational outcomes has great theoretical advantages. With the advocacy of "schooling is not learning" (Pritchett, 2013), "(student) cognitive skills" have emerged and entered the field of vision of researchers. Human capital theory holds that a person receives education in order to acquire skills, which can improve personal labor productivity, thereby contributing to personal income and national economic growth. Therefore, cognitive skills are introduced into the measurement of human capital, which is considered to be a return to the concept of human capital theory (Hanushek & Woessmann , 2015 ; Huang Bin et al., 2024). Student cognitive skills were first used by many research institutes. This is because international and regional organizations have widely carried out student ability assessment test projects in different countries. These test projects measure various aspects of literacy such as mathematics, literature , and science. The scores of each literacy can well measure students' cognitive skills. Given that the test questions used in different test projects are quite different, in order to obtain a larger data sample , it is necessary to achieve comparable conversion between different students' cognitive skill test scores. Many scholars are committed to solving this problem, and some very valuable methods have emerged (Hanushek & Woessmann , 2012 ; Angrist et al., 2021; Gust et al., 2024).

Although students' cognitive skills are increasingly valued as an indicator of education quality, and related studies have found that when both the mean of students' cognitive skills and the average years of education are controlled, only the mean of students' cognitive skills significantly affects a country's economic growth [7], the average years of education has not been completely eliminated as a measure of the quantity of education . On the one hand, this is because the years of education have an incomparable advantage in being widely accepted ; on the other hand, it is because the academic community has never denied the importance of the quantity of education and believes that the quantity and quality of education are equally important. In this context, a series of studies have emerged, which intend to combine the quantity and quality of

[7] Some scholars believe that this is due to the huge collinearity between the average years of education and the mean cognitive skills of students, rather than the fact that the amount of education is useless (Hanushek & Woessmann , 2012).

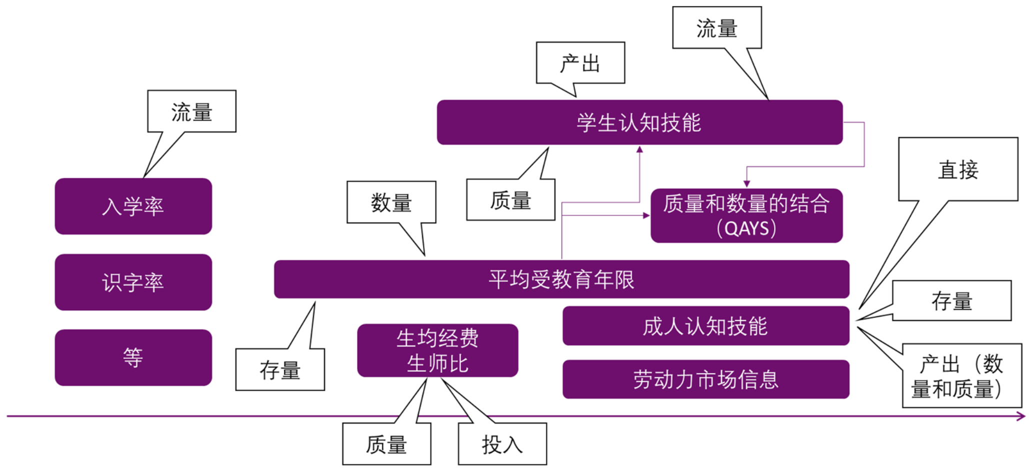
education through certain methods and construct new indicators that contain both the quantity and quality of education information. Such studies usually estimate the quality of education from students' cognitive skills data, refer to a certain benchmark (country), calculate the quality adjustment coefficient, adjust the average years of education, and obtain the quality-adjusted years of education (Filmer et al., 2020).

At the same time, the excessive use of student cognitive skills has fallen into the same defect of replacing stock with flow as the enrollment rate; and in theory, the current stock of human capital in a country should be affected by the quality of education in the student era corresponding to [8]the population, and will not be affected by the current quality of education in the country. A series of problems caused by the use of student cognitive skills can be solved by directly using adult cognitive skills data. In addition, adult cognitive skills data has both excellent qualities of educational output and stock, and according to the concept of human capital, measuring cognitive skills can be regarded as a direct measurement of human capital, so adult cognitive skills naturally become the best choice for human capital measurement indicators . However, the scarcity of adult cognitive skills test items and the limited number of countries included, in order to break through this limitation, some studies have attempted to construct adult cognitive skills data covering more countries by establishing a correlation between student cognitive skills and adult cognitive skills based on the relationship between flow and stock. This method is also groundbreaking (Egert et al., 2024) . Although this method is still in its infancy, it has great application value in the future and is worthy of attention .

In addition, some scholars have adopted different approaches to construct new human capital measurement indicators. For example, some studies have taken a new approach and completely escaped the analytical framework of education input-output. Based on the fact that wage information in the labor market is an intuitive reflection of human capital, they separated the education quality and human capital of various countries from wage information and proposed quite innovative methods (Schoellman, 2012; Martellini et al., 2024). Although this method is currently subject to the problem of obtaining wage information in the market and has

[8] In addition to theoretical inappropriateness, empirical evidence has shown that although the characteristics of students' and adults' cognitive skills are somewhat correlated, there are significant differences in characteristics such as distribution skewness and standard deviation (Huang Bin et al., 2024). These have prompted scholars to reflect on the use of student cognitive skills data instead of adult cognitive skills data in research.

shortcomings in terms of changes in education quality and human capital and measuring increments, its value cannot be ignored.



Note: Time progresses gradually from left to right along the coordinate axis.

Figure 2-1 The general development of human capital measurement indicators

Next, this article will follow the development context of national human capital measurement indicators and systematically introduce their existing problems, data construction principles and methods, existing public data, etc., to lay the foundation for the subsequent recommendation of a unified framework, comparative analysis of the advantages and disadvantages of various educational human capital measurement indicators, and summary and induction of the basic principles for judging the advantages and disadvantages of educational human capital.

III. Years of education: a [9]measure of the amount of education received

So far, the number of years of education is still the most popular human capital indicator, and the average number of years of education is still the most commonly used indicator for measuring a country's human capital in macroeconomic research. The academic community generally believes that the average number of years of education can measure the amount of education a country has obtained.

[9] It should be noted that in English, educational attainment usually refers to the level of education or the number of years of education. Its connotation itself tends to be quantitative and does not involve quality too much.

3.1 Supplementing missing data: interpolation of years of education

Generally speaking, the average years of education in each country is calculated based on individual micro data from the country's census or survey. There are two problems with the calculation of this indicator:

One is the inconsistency in the classification of education levels and types in different countries. On the one hand, the education systems and school systems of different countries are different. Each country has an education system that is different from other countries, including different types of schools, and the number of years required to complete seemingly the same level of education in different countries may vary; on the other hand, even in the same country, its school system may change at different times. To solve this problem, it is necessary to make a detailed comparative analysis of the education system and school system of each country in the sample. This work does not involve complex technology, but because it is tedious and complicated, it takes a lot of time and effort to organize.

Second, the population census of each country is usually conducted every 10 years, which is too long. If we want to form a panel data of years of education with a shorter interval (such as once every 5 years), we need to interpolate the data for the intervening years without a population census or survey. In order to achieve data interpolation, scholars have proposed several methods. The methods proposed in early studies are relatively simple, including the perpetual inventory method (Barro & Lee , 1993 ; Barro & Lee, 2001), simple linear interpolation (De La Fuente & Doménech , 2000 ; De La Fuente & Doménech , 2006) , etc. [10]In 2007, researchers began to use population birth cohort information for data interpolation and proposed the birth cohort trend extrapolation method (F orward and Backward Extrapolation) ([11]Cohen & Soto, 2007; Barro & Lee, 2013; De La Fuente & Doménech , 2015; Barro & Lee, 2015) and the birth cohort iterative backward extrapolation method (Iterative Multi-dimensional Cohort-component Reconstruction) (Lutz et al., 2007; Bauer et al., 2012; Goujon et al., 2016; Speringer et al., 2019).

The most commonly used database in international macroeconomic research today

[10] In comparison, the principle of simple linear interpolation is relatively simple, so this article will not introduce this method. For earlier interpolation methods for years of education, see De La Fuente & Doménech (2006).

[11] This series of articles did not name its methods, we extracted keywords as the names of its methods.

is based on the method of Barro & Lee (2013) , which uses the birth cohort trend extrapolation method. In order to enhance the understanding of the average years of education, and taking into account the development history and applicability of the method, this article will introduce the mainstream (formerly) perpetual inventory method, birth cohort trend extrapolation method, and birth cohort iterative backcasting method. The principles of these three methods are different, but they are related to each other: the perpetual inventory method uses the base period data to interpolate the data of subsequent years, the iterative backcasting method uses the latest data as the base period to interpolate the data of previous years, and the trend extrapolation method is based on the base period data. Both forward and backward push.

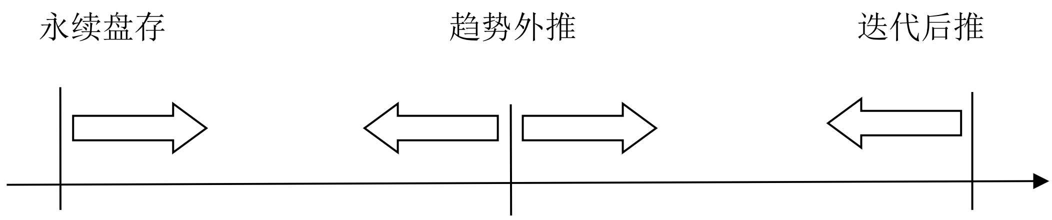


Figure 3-1 Schematic diagram of three data interpolation methods

3.2 Interpolation method for years of education

3.2.1 Perpetual Inventory Method

The perpetual inventory method uses the census or survey data of a country in a certain period as the base period, and takes into account the changes in the population at each educational stage in subsequent periods due to deaths and new enrollments [12], thereby supplementing the data on years of education in subsequent years.

Generally speaking, the average years of education of the population of each

[12] For the sake of statistical integrity, there are two statistical methods. The first one is commonly used, and is also used in this article and the cited literature. It divides all the population vertically according to the highest level of education. At this time, the population at each higher education stage is the population with the highest education level reaching that education stage (such as the uneducated population, the population with primary school education, the population with junior high school education, etc.), and the corresponding years of education are all the time required to obtain the corresponding education stage, such as 6 years for primary school, 9 years for junior high school, 12 years for high school, etc.; the second one is to stratify the population horizontally. At this time, the population at each higher education stage is the population that has completed that education stage (such as the uneducated population, the population who went to primary school, the population who went to primary school and then to junior high school, the population who went to junior high school and then to college, etc.), and the corresponding years of education are the time required to complete that education stage, such as 6 years for primary school, 3 years for junior high school, 3 years for high school, etc. In fact, after obtaining the educated population structure of each country, both statistical methods can obtain consistent results.

country is obtained by multiplying the proportion of the population at each education stage by the time required to obtain that education stage and adding them together [13], that is:

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

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

Among them, $ys_t$is $t$the average years of education of the population of a country in the year; $l$represents each stage of education, usually using a broader education level division: when no education is received $l = noedu$, when primary education is completed $l = pri$, when secondary $l = sec$ education is completed, when higher education is completed $l = ter$; $h_{l,t}$represents the proportion of the population in each stage of education, which is equal to $l$the number of people in each stage of education ($H_{l,t}$) divided by the total population (L_t); $Dur_{l,t}$is the number of years of education required to reach this stage of education in the country.

Assume that we have the population of each education stage in the base period ( ), and we need $H_{l,t-5}$to supplement the population of each education stage ( ) five years later . According to the perpetual inventory method, the following formula can be used:$H_{l,t}$

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

Where $\delta_t$is the mortality rate, $H_{l,t-5}(1 - \delta_t)$is the surviving population at each education stage after adjustment for death , $add_{l,t-5\ to\ t}$ is the change in the new population at each education stage, which is usually calculated using enrollment rate information.

This formula can be intuitively understood as the population at each education stage equals the population that completed that education stage in the base period minus the number of deaths at that education stage plus the number of new people who completed that education stage. For example, a country conducted a population census in 2010, so the average years of education of the population in 2010 can be calculated

[13] The existing data on average years of education are usually constructed based on the population aged 15-64. For the sake of simplicity, age will no longer be emphasized in the following text.

based on the above formulas (3-1)-(3-2). However, if we need the data on the years of education of the population in 2015, we can apply formula (3-3), that is, calculate the following formula:

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

As can be seen from the above formula, the mortality rate does not change with birth cohort (age) and education stage, which is considered to be the biggest problem of the perpetual inventory method, because the reality is often that the higher the education level, the lower the mortality rate (Balaj et al., 2024); the older the age, the higher the mortality rate [15]. Some scholars suspect that this defect may cause a large construction error, which is probably the main reason why relevant empirical studies cannot obtain satisfactory results (Krueger & Lindahl , 2001 ; De La Fuente & Doménech , 2006).

3.2.2 Birth cohort trend extrapolation

Using birth cohort information can effectively reduce the measurement error caused by mortality assumptions (Cohen & Soto , 2007). From the birth cohort perspective, the formula for the average years of education of the population is as follows:

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

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

Where $ys_t$is the final total average years of education, $ys_t^a$is the average years of education of each birth cohort, $l_t^a$is the population share of each birth cohort, and $h_{l,t}^a$is the population share of each education stage in each birth cohort. The new symbol introduced here $a$is birth cohort. There are slight differences in the birth cohorts used in various studies, but usually $a = 1$is 15-19 years old, $a = 2$is 20-24 years old, and so

[14] In the application of this method, the most important thing is to estimate the mortality rate in each period $\delta_t$and use the enrollment rate information to calculate the changes in the new population in each educational stage $add_{l,t-5\ to\ t}$. Due to space limitations, this article will not introduce this in detail. If necessary, please refer to Barro & Lee (2001) or contact the author of this article to obtain the specific details and explanations of these literature methods summarized by the author of this article.

[15] In addition, the occurrence of immigration will also cause changes in the proportion of the population in each educational stage of the same birth cohort at different times, but existing studies have only briefly discussed this, and no study has taken this into consideration in the methodology.

on, until $a = 1165$ years old and above. Unlike before, here the entire population is expanded into each birth cohort by age. After obtaining the average years of education of each birth cohort (ys_t^a), the weighted sum is obtained according to the population share of each birth cohort (l_t^a) to obtain the average years of education of the population (ys_t) .

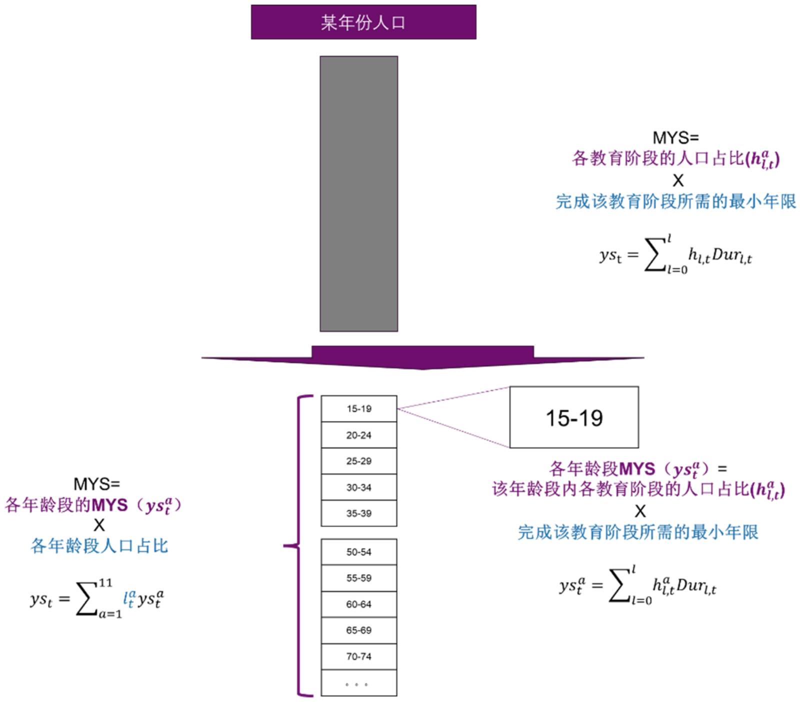


Figure 3-2 Schematic diagram of calculating average years of education

In the above formula, population information can be obtained from the World Population Prospects (WPP) of the Population Division of the Department of Economic and Social Affairs of the United Nations and can be regarded as known. Therefore, if you want to obtain the average years of education in each country, you only need to obtain one of the proportion of the population at each education stage in each birth cohort in each country ($h_{l,t}^a$) (Barro & Lee , 2013) or the average years of education (ys_t^a) (Cohen & Soto , 2007). These two pieces of information can be obtained using the birth cohort extrapolation method.

The birth cohort extrapolation method assumes that for a person who has completed formal education, the level or years of education will not change throughout his life, which means that in the same birth cohort, the proportion of the population at each education stage will remain unchanged, unless the population birth cohort undergoes major structural changes due to death (Barro & Lee , 2013). Since the number of years of education of an individual remains unchanged throughout his life, and the proportion of the population at each education stage in the same birth cohort remains unchanged, the average years of education of the population in this birth cohort will also remain unchanged (Cohen & Soto , 2007). Based on this inference, we can extrapolate the data on the number of years of education of the population in previous and subsequent years based on the base period data .

It should be noted that the birth cohort extrapolation method also makes an important assumption about the mortality rate of the population. It assumes that the mortality rate of the population in the same birth cohort will not change with the level of education, that is, in the same birth cohort, the survival rate is the same regardless of the level of education. Only by assuming this can we ensure that the proportion of the population in each education stage of the population in the same birth cohort does not change over time (Barro & Lee, 2013), and then we can conclude that the average years of education of the population in the birth cohort does not change over time (Cohen & Soto, 2007). Barro & Lee (2013) found from the existing census information that this assumption is valid for the population aged 64 and below , but not for the elderly group (people over 65 years old), and the mortality rate of the elderly group needs to be adjusted. In addition, for the young population under 25 years old, because the education of these people is still developing and changing, other methods are also needed to estimate [16].

To understand this method more intuitively, we draw Figure 3 based on Barro & Lee (2013) . In the figure, period t is the base period with data, and t+5 and t-5 are the years to be interpolated. In this figure, this method needs to complete the following two parts:

First, we can use the $t$sum $ys_t^a$of the periods with data $h_{l,t}^a$to push forward and backward to obtain the $t \pm 5$sum $ys_{t\pm5}^{a\pm1}$of the periods with missing data $h_{l,t\pm5}^{a\pm1}$. Taking the solid arrows in the figure as an example, $t$we can use the sum of the birth cohorts of the period (30-34) $h_{l,t}^4$to obtain $t-5$the sum of the birth cohorts of the period (25-

[16] Although they have the same understanding, the formulas of Cohen & Soto (2007) and Barro & Lee (2013) are slightly different in birth cohort a. In the forward projection, Cohen & Soto (2007) used 25-29 in period t to infer 30-34 in period t+5, while Barro & Lee (2013) used 20-24 in period t to infer 25-29 in period t+5; in the backward projection, Cohen & Soto (2007) used 30-34 in period t to infer 25-29 in period t-5, while Barro & Lee (2013) used 25-29 in period t to infer 20-24 in period t-5.

29) $h_{l,t-5}^{3}$and $ys_{t-5}^{3}$the sum $ys_{t+5}^{5}$of the birth cohorts of the period (35-39) by pushing backward $h_{l,t+5}^{5}$and forward $t+5$respectively ys_{t}^{4}. In general, in the forward push, we can use the following two formulas to obtain the sum of the missing $h_{l,t+5}^{a+1}$years ys_{t+5}^{a+1}:

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

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

Similarly, in the backward pass, the following two formulas can be used to obtain the missing year $h_{l,t-5}^{a-1}$or ys_{t-5}^{a-1}:

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

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

The function of the above formulas is to equate the information of the same birth cohort in different years.

Second, it is necessary to supplement the birth cohorts that cannot obtain data through trend extrapolation in other ways. For example, the gray-shaded birth cohorts (birth cohorts before the age of 25 and after the age of 64) in the figure usually consider the comprehensive impact of factors such as enrollment rate, mortality rate, and immigration in [17]the supplementation process .

[17] In almost all the literature, school enrollment and mortality rates are considered, but for immigration, there is only a discussion of its impact, not its consideration.

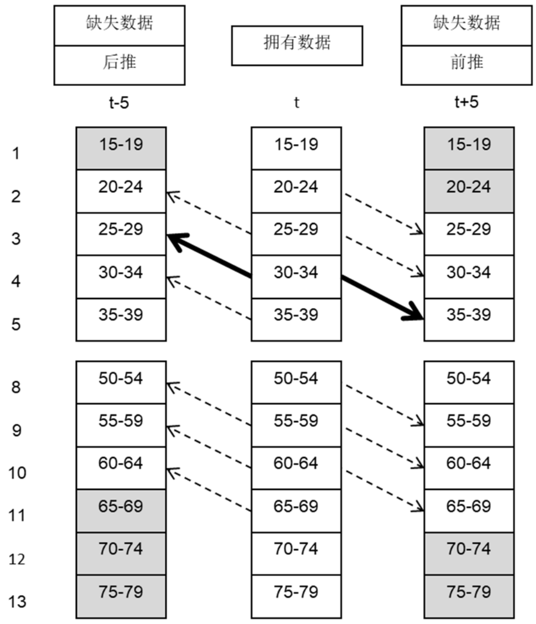


Figure 3-3 Schematic diagram of trend extrapolation method

The general steps of this type of research are usually as follows: first, the information of the birth cohort for which information is available needs to be obtained through the trend extrapolation method (i.e., formulas (3-6)-(3-8) or formulas (3-7)-(3-9)); second, certain methods are used to estimate the information of the birth cohort for which information cannot be obtained through the trend extrapolation method; finally, formulas (3-4)-(3-5) are used to calculate the average years of education in the years with missing data.

3.2.3 Birth cohort iterative backcasting method[18]

Unlike the birth cohort trend extrapolation method, the birth cohort iterative backcasting method [19] only performs backcasting, not forwardcasting. This method usually selects the latest year of data as the base year, and continuously performs iterative backcasting and value addition based on the data of that year. The implementation of this method can be roughly divided into the following two parts:

The first part is data collection and processing. The data collected in this part includes:

[18] The data constructed by this method are far less widely used than those constructed by other methods, so this article only briefly introduces this method. For more information, see Lutz et al. (2007) and Speringer et al . (2019) .

[19] The original paper called this method Iterative Multi -dimensional Cohort -component Reconstruction

(1) Education information data for the base year (t), which usually comes from the population census. After uniform data preprocessing, the population proportions of each sex, birth cohort, and education stage in the base year t are obtained $h(a, l, t, sex)$[20].

(2) The population structure data over the years is consistent with the above articles and comes from WPP.

(3) Life table data of previous years , which also comes from WPP. This data is mainly combined with the difference in life expectancy of the population of different sexes and education stages to calculate the $t-5$survival rate of the population of different sexes, birth cohorts and education stages in previous years (such as the 1990s) Survival Ratios$(a-1, l, t-5, sex)$and to adjust the survival rate.

The second step is iterative calculation. To introduce the iterative process, we take the construction of the birth cohort from 15-19 to 105+ as an example [21]. Figure 2-3 shows a schematic diagram of the complete iterative process. The text in the figure corresponds to the iterative steps and names.

[20] In this step, Lutz et al. (2007) used the population data of each sex-birth cohort-education stage, and obtained the population data through the population proportion of each sex-birth cohort-education stage and the population structure (birth cohort-population) data of the UN. Therefore, the formula in step (4) corresponds to the population, and after step (5), the number is converted into proportion and combined with the population structure data of the UN to obtain the new population data of each sex-birth cohort-education stage. This population-population proportion-population conversion is artificially considered to take into account the immigration situation. In Speringer et al . (2019) , the population proportion data of each sex, birth cohort and education stage are used from beginning to end. Only at the end, the population proportion data are combined with the UN population structure data to obtain the population quantity data.

[21]In Lutz et al. (2007) , it was 70+; in Goujon et al., (2016) , it was 100+; in Speringer et al . (2019) is 105+.

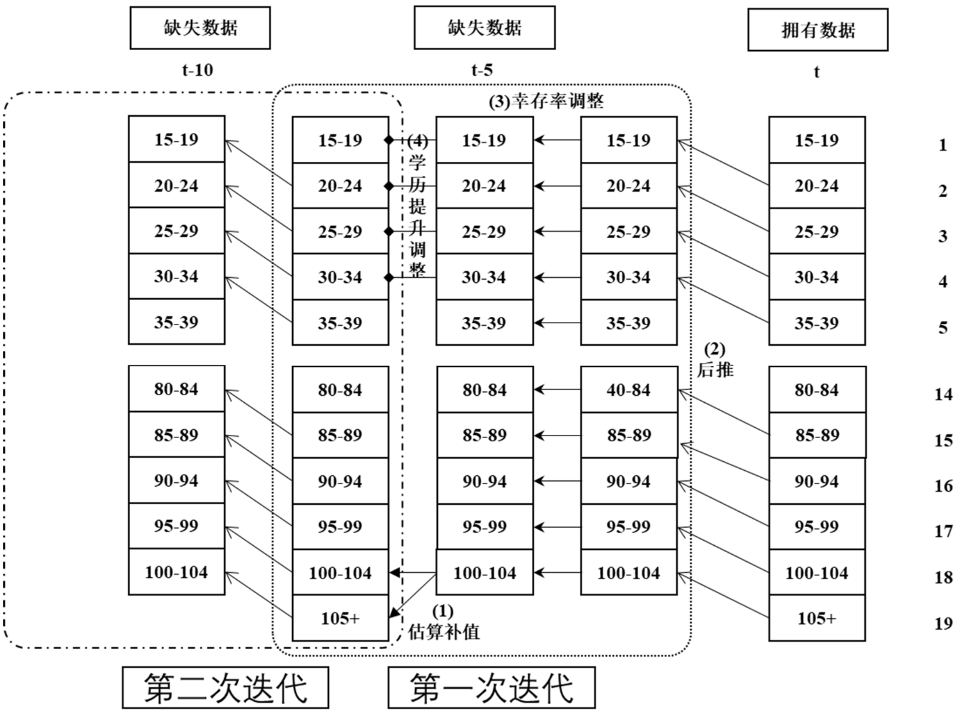


Figure 3-4 Schematic diagram of a complete iteration

In the figure, the iterative calculation has the following steps:

(1) Estimate and supplement the highest birth cohort. As shown in the figure, in the first iteration, since $t$the birth cohort 105+ of the period is used to push back $t-5$the birth cohort 100-104 of the period, $t-5$the highest birth cohort (105+) is missing for the period. If a second iteration is to be performed, this information needs to be supplemented first. Since the highest age group data will be missing only after the iteration, supplementation is required. Therefore, this step needs to be considered in the second and subsequent iterations.

(2) Backward. This step is similar to the trend extrapolation method. It only requires the use of the following formula to obtain the population proportion data of "different genders, different birth cohorts, and different education stages" for each birth cohort in period t-5:

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

This formula also equates information from the same birth cohort in different years.

(3) Make survival rate adjustments. In the process from $t$to $t-5$, death is changing the population structure, so it is necessary to use the survival rate

information Survival ratios$(a-1,l,t-5,sex)$to adjust the survival rate of the backward results:

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

(4) Adjustment for educational level improvement. Since the 15-34 year old population still has the possibility of educational level improvement, the education years of the 15-34 year old group obtained later should be adjusted to obtain the final $t-5$period result.

(5) Go back to step 1 and perform a second iterative calculation based on the results of period t-5. Repeat this process to obtain $t-10,\ t-15,\ldots$, the results of the period t-5.

The above (2) and (3) can be combined into one step. In the application of this method, various literatures basically follow the above steps to iterate and fill in the value, but the stage-by-stage processing of each part is slightly different , especially in the processing of steps (2), (3) and (5).

IV. Student cognitive skills : a measurement indicator focusing on the quality of educational acquisition[22]

Student cognitive skills are an indicator that has only appeared in the last twenty years. The mean of student cognitive skills has gradually been included in macro research. Although other statistics of student cognitive skill distribution are not the focus, they are also involved in many papers. Student cognitive skills are often used to measure the quality of education in macro research . In fact, there is little discussion on whether they can measure the quality of education. We will discuss this in Chapter 8.

4.1 Building an internationally comparable database: the comparability of scores for each test item

In the past two decades, as the academic community has recognized and

[22] The word quality has different meanings in different contexts. For example, if an education system has 5 years of education, and each year of education can improve its math score by 20 points, then the final quality of this student is 100 points, and the annual education quality is 20 points. The former is the overall quality and final quality, and the latter is the unit quality. Although students' cognitive skills are a total output, when the student's age or grade is certain, the student's cognitive skills are unit quality.

emphasized that "Schooling is not Learning", the importance of education quality has become increasingly prominent. Therefore, more and more international or regional organizations have conducted assessments on the literacy of basic education students in various countries to assess the current education quality of each country, such as the Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), Programme for Analysis of French-speaking Countries Education Systems (PESES), etc. d ' Analyse of Systems The first, second and third regional comparative and explanatory surveys (PERCE / SERCE/ TERCE) of the Latin American Laboratory for Assessment of the Quality of Education (LLECE) , the Early Grade Reading Assessment (EGRA), the Annual Status of Education Report (ASER), the United States National Assessment of Educational Progress (NAEP) , and the Latin American Laboratory for Assessment of the Quality of Education (LALECE). The National Assessment of Educational Progress (NAEP) , the India National Achievement Survey (NAS), etc. [23] These projects test various cognitive skills that can affect students' future production behaviors, such as mathematical literacy, literary literacy (reading and writing), scientific literacy, etc. The scores of these literacy and abilities can well measure the learning gains of students in various countries, and thus further evaluate the quality of the education system in various countries at that time.

In some macro studies, students' cognitive skill scores are considered the best measure of education quality [24]. However, studies usually want to have data from more countries (cross-sectional) and longer periods of time (longitudinal). However, various test items are not comparable in design. To achieve this goal, a certain method is needed to make the test items comparable horizontally and vertically.

Specifically, in the horizontal direction, the study wants to make two test items in similar years comparable [25]. There are two purposes. One is to include more countries,

---

[23] For a brief introduction to these international and regional tests, see De La Fuente & Doménech (2024).

[24] It should be noted that these studies usually do not pay attention to or ignore the time issue, and use the current education quality to regress the current economic growth. In fact, they should use the education quality of the student era corresponding to the current population in the labor market. There is a huge time gap between the two.

[25] The conversion of multiple different test items is also carried out two by two. For the sake of accuracy, two are used in the text.

especially countries at all stages of development and in all regions. However, the number of countries included in a single test item is quite limited. Even PISA, which covers the most countries, only surveyed 102 countries and regions, which is less than half of the number of countries in the world. In terms of economic development level, these countries are mostly economies with upper-middle income or above, and the coverage of lower-middle income and below economies is relatively small. In terms of regions, there are relatively few countries in Africa. The solution to this problem is mainly to make the scores of each test item comparable, so as to include more national samples by integrating international and regional test items. This is the main purpose of existing research. The second is to integrate different information between the two tests to achieve cross-regional comparison of a certain information. Since this purpose currently only appears in one type of research, we will introduce it in detail later (see 4.2.2).

In the vertical direction, macro research also usually wants to include data from more years (horizontally), which means that certain methods are needed to make the scores between multiple rounds of tests comparable. After the 1990s, each test project basically adopted the item response theory, and the same items (see 4.2.3) made the test itself comparable in the vertical direction, so no additional processing methods were needed. For the tests before the 1990s, it was equivalent to treating the tests of the same test project in different rounds as two completely different test project results. The processing method was consistent with the horizontal country processing method, but on this basis, the overall level (mean) difference that may be caused by time was considered (from this perspective, horizontal and vertical comparability are consistent). Considering the similarity of methods and the particularity and scarcity of data before the 1990s, this article will not introduce too much about temporal comparability.

4.2 Anchor point and construction conversion function

Whether expanding the national sample by integrating international and regional test items or integrating different information from two tests, it is necessary to first construct a transformation function (Transforming or Linking Function) through an anchor point and then apply the transformation function to realize the transformation of data from different test items.

Suppose you want to convert the score of test item X into the score of test item Y.

The generalized formula is:

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

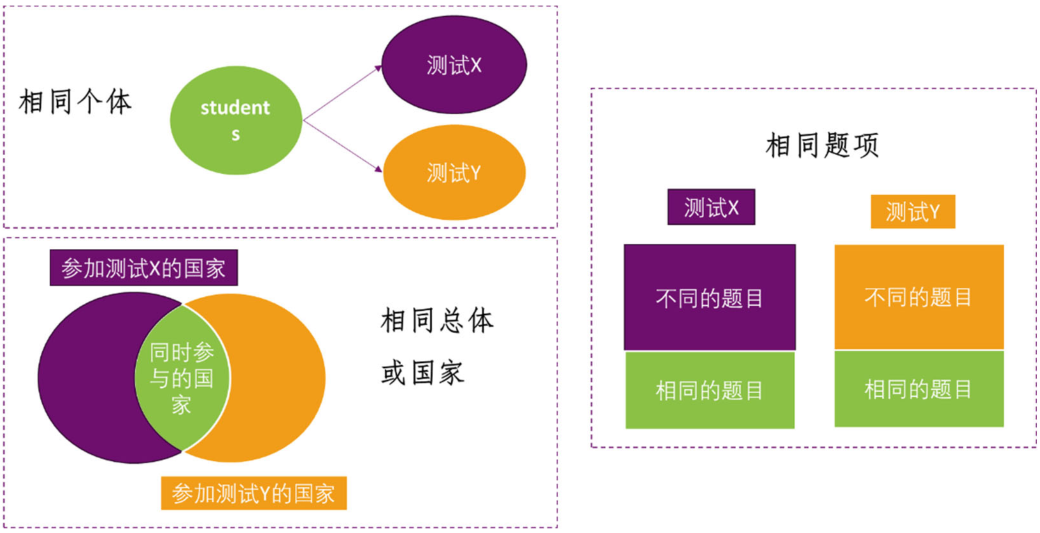
Among them, $Score_Y$is $Y$the score of the test item we want to get, $Score_X$is the score of the test item we have mastered X, $f(\cdot)$is the function, if a simple linear equation is used, the formula can be changed to:

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

Among them, $\alpha$and $\beta$are conversion parameters.

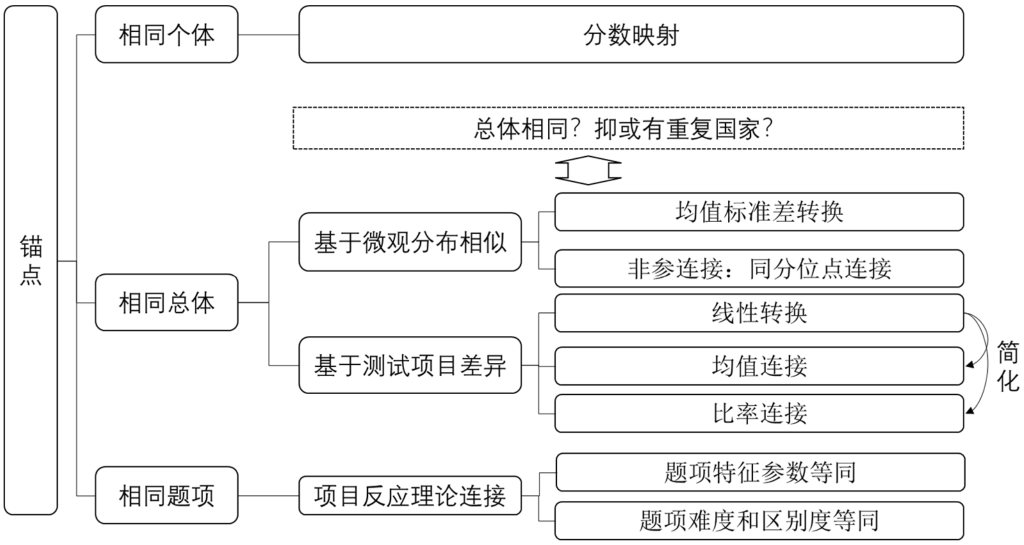
the conversion of individual student scores at the micro level , or the conversion of national average scores or total scores at the macro level, depending on the method and data level used.

The key to building a transformation function is to have anchors, which are the overlaps between two different test items. Anchors can be students, countries, and questions (or items) that overlap between two tests. Only with anchors can the transformation formula be built; it can be said that common persons, common populations, or common items or overlapping items are the basis for building transformation functions (Kolen & Brennan, 2014 ; Reardon et al., 2021) [26]. When we use different types of anchors , the techniques used will be different. The figure below gives the overall framework of the method, and we will introduce them one by one.



[26] In this article, common, overlapping, same, and repeated have the same meaning.

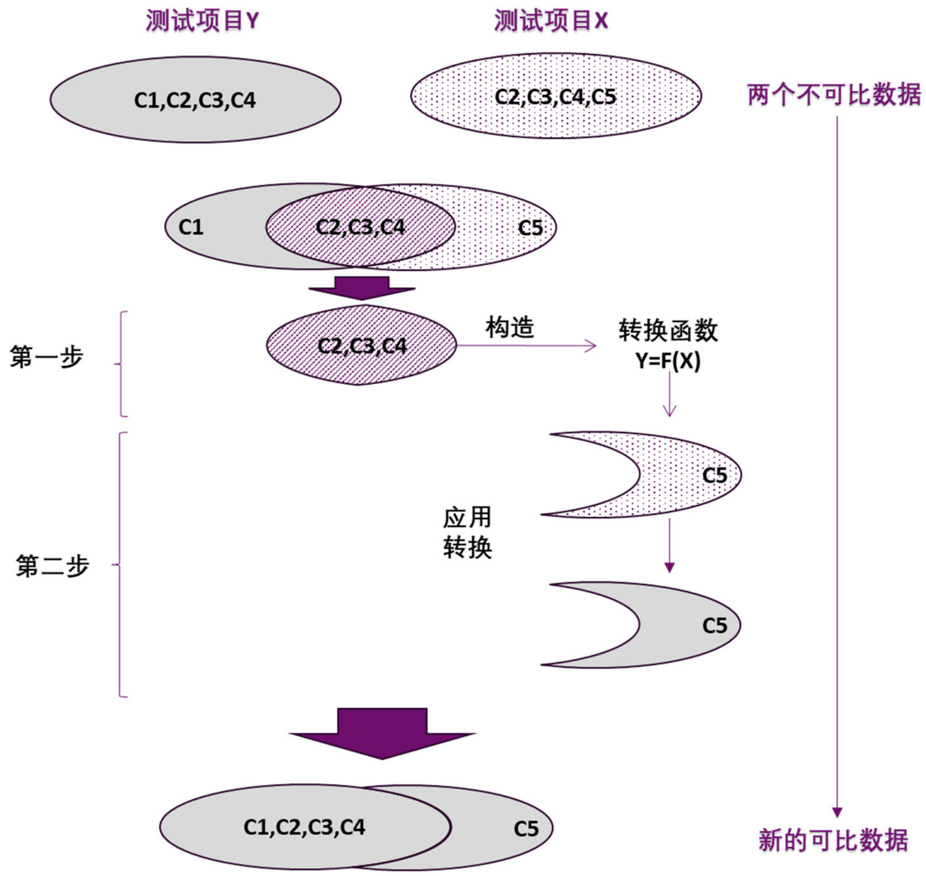
Figure 4-1 Anchor point diagram



Note: Drawn by the author.

Figure 4-2 Method framework for building conversion relationships between different tests

In addition, regardless of the purpose, as mentioned above, this type of research is usually divided into two steps: the first step is to use anchor points to build a function (tool), and the second step is to apply the function (tool) to achieve conversion. As shown in the figure below, taking the same country (see 4-2-2) as an example to expand the national sample, assume that there are two test items X and Y (indicated by different shading in the figure), both of which are tests for students' mathematical literacy, but cover different countries and regions. Test item X covers countries C1, C2, C3, and C4, and test item Y covers C2, C3, C4, and C4. Among them, C2, C3, and C4 participate in both tests (as shown in the cross-shaded part in the figure), while C1 and C2 participate in both tests. In order to expand the national sample, we must first construct a conversion function that expresses the quantitative relationship between the scores of the two tests through the national samples (C2, C3, and C4) that participate in both tests X and Y. Then, through this conversion function, the scores of countries that only participate in test item Y are converted to the scores of test item X, or the scores of countries that only participate in test item X are converted to the scores of test item Y. Taking X to Y as an example, we will get comparable data for five countries (C1, C2, C3, C4, C5).



Note: Take X to Y as an example.

Figure 4-3 Schematic diagram of research steps (taking the same country as an example)

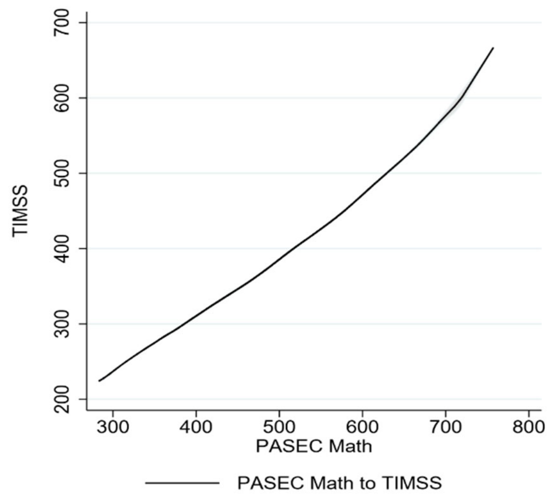
4.2.1 Using the same individual as an anchor

When two test items are taken by the same group of students and test the same abilities of the students, the differences between the test items will be directly reflected in the different scores of the same students in the two test items due to the stability of the students' abilities. At this time, the same group of students is used as an anchor point, and the local linear regression is used through the scores of the students on the two test items to directly construct the mapping function between the scores of the two different test items . It can be expressed as:

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

This mapping relationship means that Xeach score of the test item corresponds to Ya score of the test item. After obtaining the mapping relationship, it can be used for purposes such as expanding the country sample. For example, if students in a country only take test item X and do not take test item Y, we can get the score of each student in this country if they take test item Y based on the score correspondence table of X→ Y.

For example, Patel & Sandefur (2020) recruited a group of students in India and asked them to take both PASEC and TIMSS tests. Then they used local linear regression to estimate the PASEC and TIMSS test score data of these students and constructed a mapping relationship between the two test scores. As shown in the figure, it is a curve that reflects the mapping relationship between PASEC and TIMSS test scores. On this curve, there is a one-to-one correspondence between PASEC and TIMSS test scores.



Note: Based on Patel & Sandefur (2020) .

Figure 4-4 Mapping relationship represented by a two-dimensional curve

After obtaining the mapping relationship, it can be used for purposes such as expanding the country sample. For example, if $c$students in a country only take the PASEC test but not the TIMSS test, we can use the score correspondence table constructed by cPatel & Sandefur (2020) to convert the scores of students in the country into the scores of the TIMSS test, and vice versa.

individuals as anchors is that it converts individual student micro data, so the data obtained after the conversion is still micro data at the individual student level, rather than aggregated data at the national or regional level. Researchers can use student micro data to conduct more detailed analysis and achieve research intentions that cannot be achieved at the national or regional macro level. And after obtaining the mapping relationship, due to the stability of international and regional test items, that is, the comparability of each year, this mapping relationship can be used for subsequent tests.

However, this method also has some flaws. First, it is questionable whether the mapping relationship of different test scores obtained by using students from one or a few countries as anchor points can be applied to students from other countries, because this group of students may be better at one test and worse at another, resulting in deviations in the score correspondence table. Second, using individuals as anchor points to construct a mapping relationship usually requires researchers to implement a test in person, which consumes a lot of manpower, material and financial resources and has a high overall cost. Third, the mapping relationship should be able to achieve the conversion of any score, especially the high and low score intervals, so it is necessary to ensure that there are enough students in each score range to fit the mapping relationship. Fourth, in addition to this, in order to ensure the effectiveness of the mapping relationship, the entire process of test implementation, such as test question composition, student selection, and test implementation, must be strictly standardized and scientifically credible, which is undoubtedly difficult to achieve. Currently, only Patel & Sandefur (2020) in the literature uses this method.

4.2.2 Anchoring with the same population (country)

We must first distinguish between two concepts, the same population and the same country. When the participating countries of the two test projects are the same (although the students are not the same students due to sampling issues), this is the same population . However, in reality, it is often the case that only some countries participate in both test projects at the same time. In this case, the research often uses these countries participating in both tests as anchor points (to distinguish them, we call them the same countries).

To understand the difference between these two situations, let's assume that there are two test items X and Y. When both test items X and Y measure the same country C1, C2, and C3, the population formed by the three countries is used as the anchor point, and it is the same population; when test item X measures countries C1 and C2, and test item Y measures countries C2 and C3, C2 is used as the anchor point, and it is the same country.

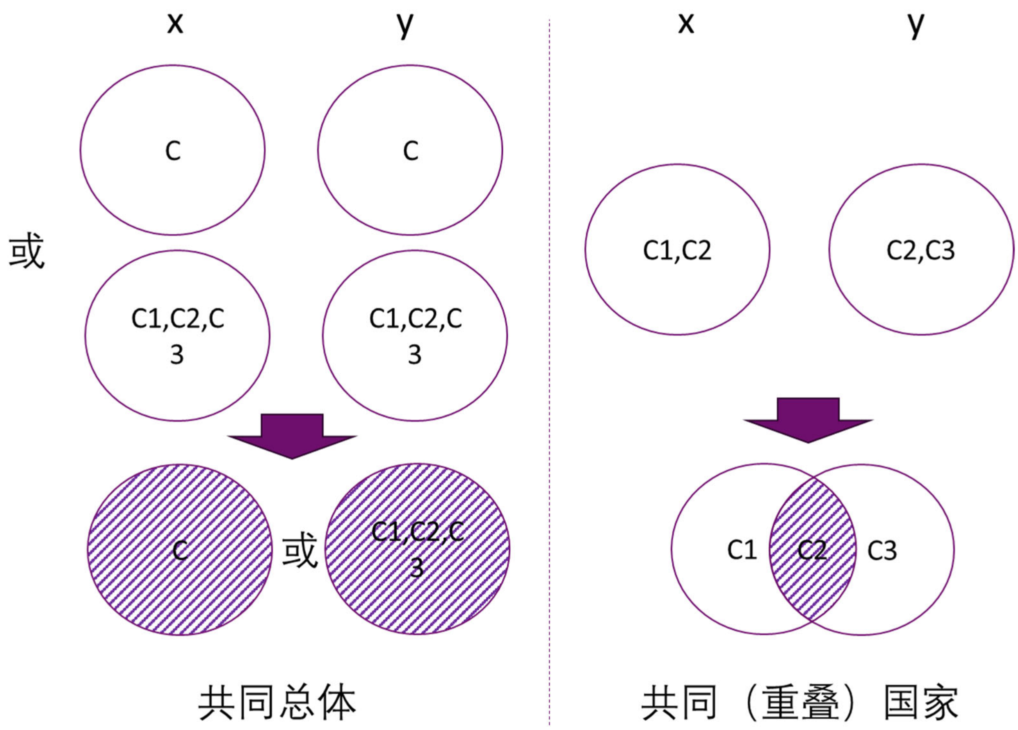


Figure 4-5 Schematic diagram of common population and common country distinction

The two situations actually reflect two different purposes. For the same population, the purpose is to integrate the information of the two tests. For example, Test X is implemented uniformly, and the scores of C1, C2, and C3 in the three countries are comparable, but there is a lack of information on the provinces (states) within the country; Test Y is implemented by each country separately, and the scores of C1, C2, and C3 in the three countries are not comparable, but each country distinguishes between the provinces (states) within the country when implementing the test, and has comparable information on the provinces (states). By treating the three countries of Test X and Test Y as the same population and integrating the information of the two tests, cross-national comparisons of provinces (states) can be achieved.

For the same country, the purpose is to expand the number of sample countries. For example, by using country C2 as the anchor country , we can use the conversion function relationship between the two test items X and Y to determine, and use this relationship to predict the score of country C1 that only participated in test item X if it participated in test item Y, and the score of country C3 that only participated in test item Y if it participated in test item X. In this way, we can expand from having comparable data for only two countries to having comparable data for three countries.

The same population (country) method needs to be based on two premise assumptions: one is the population representativeness assumption, that is, the students in each country selected to participate in the two test projects should be able to represent the same population (Linked tests must test the same underlying population) [27]; the other is the content identity assumption, that is, the skills measured by the two test projects are the same or similar (Tests should measure similar proficiencies). For example, although there are differences in design, the PISA and TIMMS tests measure students' mathematical literacy [28].

In actual research , the number of methods based on the same population (country) is the largest. These methods are mainly divided into two types according to their core assumptions :

The core assumption of the first one is the micro-distribution similarity assumption. This is an assumption about the distribution of student scores in the same population (country) in two different test projects. It believes that if different test projects can accurately measure the distribution of cognitive skills of students in different countries, then the score distribution shape of all students in the same population (country) on different test projects will be the same [29], with only differences in the mean and standard deviation of the distribution. Since this assumption is based on the distribution of micro-individual components, micro-data of individual students can be obtained . Methods based on this assumption [30]mainly include mean-standard deviation transformation and equipercentile linking , of which the latter is the only non-parametric linking method in

---

[27] In fact, since the age or grade of the subjects measured in different skill test items are not the same, this assumption is difficult to hold. In order to reduce the error caused by this problem, in the process of constructing the conversion function, we try to use similar groups, such as using the data of TIMSS eighth grade and PISA (15 years old) to construct the conversion function.
[28] What is being expressed here is that although both PISA and TIMSS have mathematics data, the difference in the questions makes their measurement biases different. Mathematics is also divided into many small items, and their small items may be different. For example, one may be more biased towards algebraic calculation ability, and the other may be more biased towards spatial geometry ability.
[29] Since the test items are measured according to the item response theory, more specifically, the score distributions of the two test items should be normal distributions. However, all methods actually only require that the two test distributions are the same, not necessarily normal.
[30] It should be noted that in a series of early studies by Hanushek (Hanushek & Kimko , 2000, Hanushek & Woessmann , 2012a, Hanushek & Woessmann , 2012b), although they were also based on the idea of mean-standard deviation transformation based on the micro-distribution similarity assumption, they used the overall mean of the country rather than the individual micro data of students. In addition, in non-Hanushek studies that used this method, the overall mean of the country was also used, such as Altinok et al. (2018). In fact, these studies treat each country as an individual in time and use the distribution formed by these national individuals. In recent studies by Hanushek, this method was applied to micro-individual data, such as Hanushek & Woessmann (2015) and Gust et al. (2024).

the same population (country) method.

The second core assumption is the test item difference assumption. This is an assumption about the difference in scores (means) of the same country. It believes that the systematic differences in scores of countries in the same country on two test items should come from the differences in the two test items, rather than the differences in countries. Unlike the first assumption, the method based on the second assumption emphasizes the overall differences, so the national mean is used to construct the conversion function. Therefore, this method obtains overall data at the national level and cannot obtain individual micro data. In addition, this type of research uses the repeated countries of the two tests as anchors, and the purpose of the research is to expand the sample countries. The methods based on this assumption mainly include Line Linking and the limited Mean Linking and Ratio Linking.

But no matter which assumption is used, the overall situation is estimated based on the same countries and using the overlapping countries' situations.

As shown in the table below , in this paper, almost all the methods in this section have been used to expand the number of sample countries, but currently only the mean standard deviation method is used to integrate the information of the two tests.

Table 4-1 Purpose - Method - Assumption Correspondence Table

Same population: integrating information	Mean Standard Deviation Transformation	Similar microscopic distribution
Same country: Expand quantity		
	Same point connection	
	Linear conversion	Test item differences
	Mean connection	
	Ratio connection	

In addition, it should be noted that the naming of the methods in this article is slightly different from that in other literature. The following table gives a comparison of the method naming.

Table 4-2 Correspondence between method names in this article and other literatures

Method naming in this article		Nomenclature in other literature
Similar microscopic	Mean Standard Deviation	Linear connection (Altinok et al., 2018; Angrist et al., 2021)

distribution	Transformation	
	Same point connection	Same point connection (Altinok et al., 2018)
Test item differences	Linear conversion	Regression (Angrist et al., 2021)
	Mean connection	Mean connection (Altinok et al., 2018)
	Ratio connection	Pseudo-linear connection (Altinok et al., 2018); ratio connection (Patrinos & Angrist , 2018)

Note: This is the author's own work.

(1) Micro-distribution similarity assumption: mean-standard deviation transformation

the first micro-distribution similarity assumption, since the shapes of the distributions of the two test items are the same, and they only differ in the mean and standard deviation of the distributions, we only need to use the conversion formula between the distributions and adjust the mean and standard deviation of the two distributions [31] to construct the following conversion function:

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

where $Score_{Yi}$ and $Score_{Xi}$ are the individual i scores [32] on test items X and respectively Y; $\mu_X$and are the mean and standard deviation of the scores of students $\sigma_X$in the anchor country on test item X, $\mu_Y$and $\sigma_Y$are the mean and standard deviation of the scores of students in the anchor country on test item Y.

At present, the only one that can be completely considered as the same population is Reardon et al. (2021) , whose intention is to compare the scores of school districts between states. However, although the national unified NAEP test project is comparable between states, it does not have school district information; at the same time, although the test projects organized by each state are not comparable between states, they have score information for each school district in the state. In this case, the same population is the states that participate in both the national test project and their own test project. By integrating the information of the two tests, cross-state comparison of school

[31] This formula is similar to standardizing a distribution and then denormalizing it, but standardization only adjusts the scale of the distribution and does not change the original information of the data.

[32] By transforming the formula, we can obtain: $Score_{Yi} = \frac{\sigma_Y}{\sigma_X} * Score_{Xi} + \mu_Y - \frac{\mu_X}{\sigma_X}\mu_Y$, so this method is also called linear transformation by Altinok et al. (2018). Since this method actually uses the micro-distribution similarity assumption, we call it mean standard difference transformation, rather than linear transformation. Even when this method is applied to macro-national level data, such as Hanusheck 's series of studies, it is still based on distribution, but it changes from the distribution of micro-individuals to the analysis of macro-individuals.

districts can be achieved.

In most studies, the same countries are usually used: in reality, it is often the case that only some countries participate in both test projects (in this case, we call them the same countries). For example, in international student test projects, some countries usually participate in TIMSS at the same time; some countries only participate in TIMSS but not PISA; some countries only participate in PISA but not TIMSS. In the absence of a strictly identical population, the study can only use as many of the same countries as possible in the two test projects Cas anchors, and use the mean and standard deviation of the score distribution of all students in these same countries to infer the mean and standard deviation of the overall student distribution (Gust et al., 2024), and transform the conversion function:

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

Among them μ_X^{pooled}, σ_X^{pooled}, μ_Y^{pooled}, and $\sigma_Y^{pooled}$are the means and standard deviations of the samples composed of individual middle school students in the repeated countries in the two test items; $Score_{PISA,i}$and $Score_{TIMSS,i}$are the scores of students $i$in PISA and the corresponding scores that can be obtained in TIMSS. In terms of procedures, such studies are usually divided into two steps. The first step is to construct a conversion function using repeated countries, and the second step is to use the conversion function to convert the student scores of countries that only participate in a certain test item into the scores of the target test.

As can be seen from the formula, the above conversion is actually composed of two conversions: level conversion (or level adjustment) by the mean and difference conversion (or difference adjustment) by the standard deviation. These two conversions correspond to the difficulty (D [33]ifficulty) and discrimination (D iscriminati on) in the item response theory (IRT). Level conversion solves the problem that the same person scores differently in different test items due to the different difficulty of the two test items, and it is assumed here that the difference caused by the difficulty is the same for all individuals (when the standard deviation is the same, this difference is $\mu_X - \mu_Y$); Difference conversion solves the problem that two identical people score differently in

[33] The content of item response theory can be found in the same item section.

different test items due to the different discrimination of the two test items, for example, two people who differ by 10 points in one test item only differ by 2 points in another test item.

Since the distribution shapes of the two test items are consistent, the students' positions in the overall will remain unchanged after the conversion. However, due to the differences in the content of the two test items, the assumption of identical content is not strictly met. For example, TIMSS focuses on school curriculum content, while PISA focuses on real-life problems (Hanushek & Woessmann , 2012a), which is reflected in the differences in test questions , resulting in certain differences in the mathematical literacy of the two tests in terms of connotation. In this case, the distribution shapes of student scores in the two test items will be different. At this time, after using the distribution transformation, the position of individual students in the distribution will change slightly, which is often unavoidable, but the large correlation between the tests (such as the math test scores of TIMSS and PISA) makes the error caused by the non-strict satisfaction of the assumption of identical content within an acceptable range. Therefore, although this method can obtain micro data of individual students, research usually shifts the focus to the distribution mean rather than the score of a particular student to reduce the impact of this error; if other information of the micro distribution of individual students after the transformation is used, caution should be exercised.

In addition, under strict data constraints, there may be only one country in common for both test items. In this case, only one country can be used as the "anchor point" for conversion (Angrist et al., 2021). The essence of mean and standard deviation conversion is to use the distribution of individual students in the anchor country to infer the distribution of the potential population. Obviously, using a single country's student sample to infer the distribution of the overall student sample score is undoubtedly less reliable. Therefore, using one country as an anchor point for conversion will undoubtedly result in a large error [34]; in fact, if the number of countries used for connection is small, the credibility of the conversion results using this method will

[34] Due to the difference in the content of the two tests, it is possible that some students in a country are better at test X, while other students are better at test Y, making the distribution shapes of the two tests in the same country inconsistent. At this time, only by using a large number of individuals from more countries can the two distributions be closer to normal distributions with similar shapes.

become questionable (Gust et al. 2024).

(2) Micro-distribution similarity assumption: Equipercentile Linking

equivalent percentile linking is a common nonparametric method for comparing test item scores that does not use item response theory (Kolen & Brennan, 2014). This method was developed by Braun & Holland (1982) .

Under the assumption of similar micro-distribution, the shape of the score distribution of all students in the same population (country) on different test items will be the same, so the two distributions can be directly connected using the same quantile points.

This approach typically consists of two elements:

One is the same quantile point value. That is, according to the quantile of the score distribution, the corresponding scores are connected, which is more intuitive as shown in the following figure:

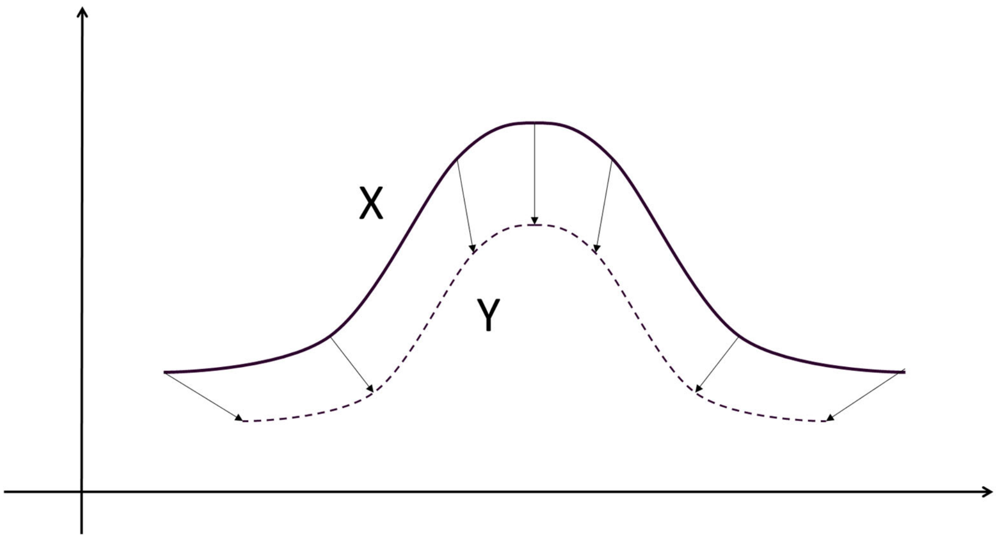


Figure 4 -6 Schematic diagram of the same quantile connection

There are two score distributions X and Y in the figure. The value (i.e. score) of each quantile (12th percentile) of X corresponds to the value of the same quantile (12th percentile) of Y. If expressed in a formula, it can be written as follows:

$$Q_X^q = Q_Y^q (4\text{-}6)$$

The second is smoothing. In the same quantile connection, because the scores are relatively discrete, we may not be able to find the score and quantile score of another distribution corresponding to a certain score or a certain quantile score, so smoothing is required. For example, X's 100 points and 105 points correspond to Y's 103 points and 110 points respectively, but because the score values are relatively discrete, it is unknown how many points in Y correspond to X's 102 points; for another example, in an ideal state, X's 47th percentile should correspond to Y's 47th percentile score, but because the score values are discrete, the closest scores in practice may correspond to X's 47.2th percentile and Y's 47.6th percentile. At this time, although we can obtain a rough match, it is not accurate enough.

In this case, although we can use percentile rankings , the accuracy is insufficient; we can also increase the sample size to alleviate this problem, but it is often insufficient. For this reason, smoothing methods have been developed to deal with sampling errors and obtain empirical distributions and quantile connections that are closest to the underlying population (Altinok et al., 2018).

According to the two smoothing methods, the same quantile connection can be divided into the following two types: Pre - smoothed E quipercentile Linking and Postsmoothing E quipercentile Linking). In pre-smoothing homoquantile joining (smoothing first, then equalizing) , the score distribution is usually smoothed first using polynomial loglinear , and then a corresponding table is constructed for the quantile values (Holland and Thayer, 2000). In post-smoothing homoquantile joining (equalizing first, then smoothing), the score distribution is first used for equalizing the quantiles, and then a cubic spline is used for smoothing (Kolen, 1984).

In fact, since it is based on quantiles, this method does not strictly require the distribution shapes of the two test items to be consistent, but requires that the distribution positions of the same individuals in the two test items are the same. This is equivalent to relaxing the assumption. Therefore, this method is most suitable for the difficulty of the two test items to change nonlinearly (Altinok et al. , 2018).

As before , the goal is usually to convert individual scores of students from countries that only participated in one test project into scores of the other test project based on countries that participated in both test projects. Therefore, the basic approach of this method is to use the score distribution of students in the same country to infer the score distribution of the whole, then build a corresponding relationship between the distributions of the two test projects, and finally transform the student-level data of countries that only participated in one test project. And, similarly, the number of the same countries is an important factor affecting the quality of the connection. Because only when there are enough countries and enough student samples, the distribution of the two tests can be close to the distribution of the whole; there are also enough samples to estimate different quantiles. The literature using this method is mainly Altinok et al. , (2018), Sandefur (2018) .

(3) Testing the Item Difference Hypothesis : Linear Transforming

In the same country, the systematic differences in the scores of different countries in the two test items should come from the differences between the two test items, rather than the differences between countries. Therefore, the conversion parameters can be obtained directly by using OLS to estimate the following formula :

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

Among them, $\mu_{X,c}$and are $\mu_{Y,c}$the mean scores of $\alpha$the anchor countries $c$in the test $X$and test , Yrespectively , and $\beta$are used to capture the systematic differences between the two test items [35]. Like the mean standard deviation transformation, $\alpha$and βcorrespond approximately to the level adjustment and difference adjustment in the distribution transformation, respectively, reflecting the difficulty and discrimination differences of the two test items. The in the model $c$represents the countries that participated in two different test items in the same round of test items. This method also uses common countries to estimate the population. After constructing the conversion function using the same countries , the total score of the countries that only participated in the X test can be converted to the total score under the Y test.

It should be noted that after the 1990s, the tests were made comparable across rounds using the same items (see below), which also means that the difference between two test items will remain the same across rounds. Therefore, the above formula can be estimated using multiple rounds of data ($r$denoted as round):

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

[35] This method is inconsistent with the linear transformation in $\frac{cov(X,Y)}{var(X)}$Altinok et al. (2018), because the beta estimation coefficient of this method is , while in Altinok et al. (2018) $\frac{\sigma(Y)}{\sigma(X)}$, the relationship between the two is; $\beta = \rho * \frac{\sigma(Y)}{\sigma(X)}$, where $\rho$is the correlation coefficient. However, it is undeniable that the mean connection and ratio connection can also be derived from the mean standard deviation transformation, which is also the order of the text in Altinok et al. (2018).

The method of estimating conversion parameters by regression has only been used recently (Angrist et al. , 2021). Although this method is also limited by the number of samples in common countries, as time goes by and the test project continues, more and more samples will be available for estimation, and the accuracy of the estimated conversion coefficient will become higher and higher. However, the problem with this method is that with each additional round of testing, the parameters estimated by the method will be slightly different.

(4) Testing the Item Difference Hypothesis : Mean Linking

If the same country participates in two test items (X, Y) and obtains two scores ($Score_X$and $Score_Y$), the relationship between the two scores can be given by the following formula:

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

Based on this idea, mean connection assumes that the difference between two test items is a fixed constant: α. That is, the difference between the two tests will be given by:

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

The constant αcan be estimated using the same countries participating in both test programs [36]:

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

Among them, $\mu(Y)$and $\mu(X)$is the average of the means of the same country in the two test items. Due to the limitation of vision, the author has not found any literature that uses this method for test items after the 1990s, but this method, like linear transformation, can use multiple years of data to estimate the conversion coefficient when used for the conversion of test items after the 1990s.

Compared with linear transformation, this method limits the conversion function to a simple addition and subtraction form, which actually limits the form of the difference between the two test items. If the form of the actual test item difference is not the same, a large error will occur. Only when the discrimination of the two test items

[36]We can also use regression to estimate directly by limiting the slope to 1.

is consistent , the conversion result of this method is consistent with the linear transformation.

This method was first introduced in Altinok et al. (2018), but was already used in Hanushek & Kimko (2000) .

(5) Testing the Item Difference Hypothesis : Ratio Linking

This method is mainly derived from a series of abstracts of the World Bank (Altinok & Murseli , 2007 ; Altinok et al., 2014 ; Altinok et al., 2018 ; Patrinos & Angrist, 2018). If the same country participates in two test projects (X, Y), and obtains two scores ($Score_X$and $Score_Y$), the relationship between the two scores can be given by the following formula:

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

Based on this idea, the ratio connection considers the difference between two tests to be a fixed ratio: β. That is, the difference between the two tests will be given by:

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

βIt can be estimated using the same countries participating in both [37]:

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

where $\mu(Y)$and $\mu(X)$is the average of the means of the same country in the two test items. This method is the same as the linear transformation. When used for the conversion of test items after the 1990s, multiple years of data can be used to estimate the conversion coefficient.

Compared with linear conversion, this method limits the conversion function to a simple product, which actually limits the form of the difference between the two test items. If the form of the actual test item difference is not the same, a large error will occur. Only when the difficulty of the two test items is the same, the conversion result of this method is consistent with the linear conversion.

(6) Summary

No method can convert scores perfectly. What is important is that the impact of non-test items such as different learning progress in a certain country is small enough. The methods listed in the article have adopted certain methods to reduce the impact of

[37] It is also possible to limit the intercept term to 0 and use regression to estimate directly.

these factors, such as " mean and standard deviation transformation" and "isoquantile linking" based on the calculation of transformation parameters and links after pooling individual samples of students in the anchor country ; the residual term in linear transformation, mean linking method (Mean Linking) and ratio linking method (Ratio Linking) use the mean average of $\mu(X)$the anchor country in the two test items $\mu(Y)$.

4.2.3 Using the same item as an anchor : IRT Linking

It is conceivable that if the two test items contain exactly the same items , then the scores of different student groups are directly comparable. However, this ideal is not realistic. Even the same test will be different in different rounds. In fact, if we lower the requirements and only need a certain amount of overlapping items in the two test items, we can use these repeated items as anchors and achieve comparability of the scores of the two test items according to certain methods.

of the same items is that both test projects rely on Item Response Theory (IRT). As a modern psychometric theory, IRT is widely used in international and regional test projects. IRT holds that the probability of a student correctly answering a given test item is a function of the student's characteristics and the test item's characteristics .

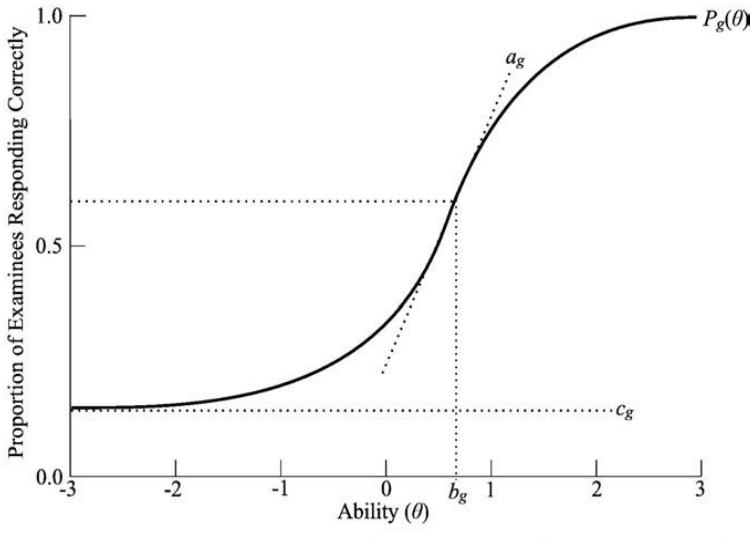
Specifically, in one of the most commonly used three-parameter logistic models (3PL), for a binary answer question ($X_{ig} \in \{0,1\}$, where 0 indicates an incorrect answer and 1 indicates a correct answer), the item response function (IRF) gives the probability that an individual with literacy $\theta_i$has ia difficulty (D ifficulty) of b_g, a discrimination (D iscriminati on) of a_g, and a probability of guessing correctly (Pro robability of Guessing Correctly , also known as guessing degree) of c_ga [38]question g:

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

Among them, θit is also called latent variable, which is usually a variety of thinking, ability, literacy, characteristics, etc. In international and regional student tests, it is usually mathematical literacy, reading literacy and scientific literacy. The figure below is the item response curve (IRC) of the three-parameter model, which gives an

[38] The scores of test items using IRT are mainly determined by difficulty and discrimination, which also determines that the conversion connection between tests should adjust the difficulty and discrimination.

intuitive presentation of the role of the parameters.



Note: Cited from Das & Zajonc (2008).

Figure 4-7 Item Response Curve

Depending on the parameters, the above formula can be simplified into a single-parameter model (difficulty) and a two-parameter model (difficulty, discrimination). Knowing whether the student answered correctly and the characteristic parameters of the item , the student characteristic parameters (i.e., the student's ability and literacy) are estimated according to the model to obtain the student's test item score. There are two key assumptions in item response theory: the unidimensionality assumption of the measured latent variable and the parameter invariance assumption. The unidimensionality assumption means that all items that make up a test measure the same latent trait; the parameter invariance assumption means that the characteristic parameters of the question are fixed for any population and are not affected by the distribution of the examinee's ability.

In principle, if two test items have a certain number of identical items, the two test items can be connected based on the identical items to achieve comparability of the scores of the two test items. For example, if we take a test item that has three rounds and each round has only two questions, the questions in each round are (question 1, question 2) (question 2, question 3) (question 3, question 4), and the two consecutive rounds have one identical question: (question 2) and (question 3). When the characteristic parameters of question 1 are fixed and known, since both question 1 and question 2 should estimate the same ability (latent variable), the characteristic parameters of question 2 can be determined. Similarly, given the characteristic parameters of question 2, the characteristic parameters of question 3 can be estimated. Similarly, the third and second round test items can be linked to the first round test items to achieve comparability of multiple round test items (Das & Zajonc , 2010) . It should be noted that although the above example uses a repeated item , in actual practice, the two test items should have a certain proportion of repeated items to achieve the purpose of reducing linking errors (Hastedt & Desa, 2015).

In fact, after the 1990s, international and regional test projects would set a certain number of repeated items in different rounds of test items, so that the scores of each round of test items can be directly compared over time (Angrist et al., 2021) [39].

Different methods will be generated depending on the setting of item feature parameters . Here we mainly introduce the following two ways to build connections (Sandefur , 2018):

(1) Item feature parameters are equal

In this case, the same items in the target test (usually a regional test, such as SACMEQ) are used, and the characteristic parameters [40]of these same items in the reference test (usually an international test, such as TIMSS) are used .

That is:

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

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

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

The subscripts $r$and $t$represent reference test items (Reference Tests) and target test items (Target Tests) , respectively .

Literature using this approach includes the examples above, as well as Das & Zajonc (2010), Singh (2014), and Sandefur (2018) .

[39] For studies that convert test scores before 2000, see Hanushek & Woessmann (2012a) , where the anchor point is a test that can be compared across time, typically the NAEP test in the United States.

[40] The validity of this method relies on Differential I tem Functioning (DIF) to test.

(2) Mean -sigma Linking

Unlike item characteristic parameter equality, it does not make the characteristic parameters of the same items of the target test items and the reference test items consistent, but requires that the average difficulty and discrimination [41] of the same items remain unchanged in the two tests . Based on the invariance assumption in the core assumption of item response theory, the scores of two equivalent test items can be related by a linear transformation:

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

As before, for the same population (country), coefficient A is used to adjust discrimination and intercept B is used to adjust difficulty. Similar transformations can be applied to the characteristic parameters of the items :

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

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

In theory, only one identical item is needed to obtain the two parameters of $A_{rt}$and B_{rt}, but this approach is not feasible in practice due to measurement errors and imperfect model fit. In fact, researchers are forced to choose between the different sums Bobtained for each item .A

Or take a simpler method, called the mean-sigma method, which uses the mean and standard deviation of Adifferent characteristic parameters $b$of the same item to obtain the sum B:

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

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

After obtaining the conversion coefficient, it can be directly applied to the student's literacy parameters $\theta_i$to convert the target test item score into the reference target score. In other words, this method can be regarded as estimating the parameters of the linear conversion formula in the same population (country) from the level of item

[41] The literature does not mention how to set the guessing degree.

characteristic parameters . Literature using this method includes Sandefur (2018) .

V. Integrating years of education and students’ cognitive skills: a quantitative and qualitative approach

5.1 Constructing indicators that include both quality and quantity

In the process of shifting the focus from the quantity of education to the quality of education, the academic community does not completely deny the importance of the quantity of education or abandon the average years of education, but believes that the quantity and quality of education are equally important. Therefore, some studies have attempted to combine the quality and quantity of education, and two approaches have emerged: linear combination and product combination.

Before we get into the analysis, we need to go back to a question: the average years of education can only measure the quantity of education obtained, which has become a consensus after the efforts of many scholars. However, can students' cognitive skills measure the quality of education obtained?

According to human capital theory and education production function theory, cognitive skills are an important component of human capital and are the joint output of input factors such as school education, family education, and personal talent. For school education, existing research usually simply breaks it down into two aspects: education quantity and education quality. In other words, from a theoretical point of view, the number of years of education refers to the amount of school education an individual has received, which is an indicator of education input; while students' cognitive skills are an indicator of education output, which is the joint output of education quantity and education quality, rather than a direct measure of education quality.

It seems that years of education and students' cognitive skills are two indicators of different dimensions, which cannot be simply merged together.

According to the above discussion, the correct approach should be to use the education production function to separate the quality of education from the students' cognitive skills. At the same time, when students are in the same grade or age (that is, the same amount of education), students' cognitive skills are only determined by the quality of education, and students' cognitive skills can be used to reflect the quality of education. In other words, only when students are in the same grade or age can we use students' cognitive skills to measure the quality of education; when students are not in the same grade or age, we need to use the education production function to estimate the quality of education from students' cognitive skills. [42]We will return to this point in Chapter 8.

The above analysis also tells readers that the key to using students' cognitive skills is not only the comparability of each test item, but also the consistency of students. However, when using students' cognitive skills, existing research usually prioritizes data availability and puts student consistency second: that is, other test data are used to fill in the gaps, without considering or forcibly ignoring the inconsistency of students' age and grade (Gust et al, 2024).

There are also studies that estimate the quality of education received by adults at different times based on adult cognitive skills, and calculate the quality-adjusted years of education based on the years of education received by adults (Hanushek & Zhang, 2009). Unlike student cognitive skills, using adult cognitive skills requires the use of an education production function to estimate the quality of education from adult cognitive skills.

Next, we will start from the human capital (cognitive skills) production function and give a detailed introduction to this aspect.

It is important to note that much of the following discussion assumes that school quality does not change over time or across grades, that is, the quality of education this year is the same as tomorrow; the quality of education in grade 3 is the same as in grade 4.

5.2 Human capital (cognitive skills) production function and linear and multiplicative combinations

5.2.1. Human capital (cognitive skills) production function

According to the theory of education production function, education output is the

[42] When the cognitive skills of the same grade are used as the quality of education, the research will not cause any problems because the amount of education is the same; when the cognitive skills of different grades are used as the quality of education, there will be certain errors in using students' cognitive skills as the quality of education because the amount of education is different.

final product of school education, family education, extracurricular education and other factors (Hanushek & Zhang, 2009). The impact of school education on cognitive skills is a function of education quality and education quantity. Combined with Hanushek & Woessmann (2012a) , the human capital (cognitive skills) production function can be expressed as follows:

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

Among them H are human capital, which here mainly refers to student/adult cognitive skills (CS); F family investment; $\phi(n,q)$ school investment, including the quantity (n) and quality (q) of school education [44]; A personal talent; Z and other factors, including labor market experience, health, etc.

Just like " Schooling is not learning" , " Learning is not just schooling" , however, in general, research focuses more on school education. Because compared with extracurricular education, school education has a greater impact on cognitive skills (Filmer et al., 2020), so other parts are often ignored by some studies:

$$H(CS) = f(n,q) \text{(5-2)}$$

The most important message conveyed by this production function is that receiving school education does not necessarily lead to the substantial development of personal skills and the effective accumulation of human capital. For personal skill development, the quantity and quality of school education are only inputs, and the function conversion process from school education input to personal skill development must be experienced .$f(Schooling, Quality)$

However, $f(Schooling, Quality)$the function form of is unknown and needs to be set manually. The function form can be divided into two points for discussion. One is the combination of education quantity and education quality, and the other is education production efficiency, that is, how the combination of education quantity and quality affects education output. Currently, there are two forms of the combination of education quantity and quality: linear combination and product combination .

---

[43] For simplicity, the above ignores the symbols of subjects and time.

[44]This refers to the unit mass.

(1) Linear combination

Linear combination means that the relationship between the quantity and quality of education is a linear combination, so this function is usually:

$$H = f(rS + wQ)\text{(5-3)}$$

In Hanushek et al. (2017) and Angrist et al. (2020) , the exponential function form is used:

$$H = e^{rS+wQ}\text{(5-4)}$$

Among them, s is the quantity of education, usually the average years of education; Q is the quality of education, usually the cognitive skills of students; $r$and q $w$is usually estimated from the micro-level Minser income equation.

First, it is not difficult to see that after taking the logarithm of both sides, this form is very close to the Mingser income equation, which may be the reason for taking the logarithmic form. Secondly, existing studies often use the cognitive skills of students of different grades or ages as the quality of education and put them into the regression, ignoring that the cognitive skills of students of different ages or places are not completely equivalent to the quality of education.

(2) Product combination: Quality- adjusted years of schooling (QAYS)

Product combination means that the quantity and quality of education are in product form in the function:

$$H(CS) = f(SQ)\text{(5-5)}$$

Unlike before, in order to make this product meaningful, Qhere it is usually the quality adjustment coefficient, rather than the direct education quality, so SQit can be understood as the years of education after quality adjustment [45]. The formal formula can be expressed as follows:

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

Where $QAYS_c$is the adjusted years of education, $S_c$is $c$the average years of

[45] These studies include Hanushek & Zhang (2009), Kaarsen (2014), Filmer et al. (2020), Reiter et al. (2020) , Glawe & Wagner (2022) , etc. In some studies, the quality-adjusted years of education are directly equivalent to human capital, while in other studies, a function is applied to it to calculate human capital, such as Kaarsen (2014) . Here we can also see the confusion in the use of related concepts.

education in a country, Q_c^b and is the quality adjustment coefficient. The quality adjustment coefficient is generally based on a certain country and can be calculated using the following formula:

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

Among them, q_c is the education quality of the target country, and q_b is the education quality of the reference country (B enchmark Country).

$QAYS$ In fact, the average years of education are transformed to include information on the quality dimension without abandoning it.

5.2.2. Choice of Functional Form: Different Types of Cognitive Skill Production Functions

For linear combination, the key to the problem is to estimate the coefficients r and w of the combination of quantity and quality. As mentioned earlier, this is mainly estimated by micro data.

For product combination, the key issue is to estimate the quality of education. The quality of education reflects the productivity (Filmer et al., 2020) or effectiveness (Kaarsen , 2014) of a school education system .

The estimation of education quality also needs to return to the human capital (cognitive skills) production function.

Since student/adult cognitive skills are usually national averages, the production function in product form can be rewritten as:

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

$CS_c$is the cognitive skills of students/adults in the country c, is $n_{c,l}$the number of years of education corresponding to $l$the country c's grade , $l$which is usually 4th or 8th grade in existing international student testing projects, and is the quality of education received by students of different grades in $q_c$the country c's education system . Because it is assumed that the quality of education does not change with grade, the subscript of the symbol for education quality is only the country c[46].

Even if the quantity and quality of education will enter the regression in the form of a product, the above also requires a clear form of the production function in order to estimate the quality of education, which requires further discussion on the meaning of education quality. In the existing literature, there are two different ways of understanding education quality from the perspective of input and output:

One is that the quality of education is an input, and there is a problem of production efficiency between input and output. The input can be a diminishing marginal return, such as in Kaarsen (2014), $f(\cdot)$in logarithmic form:

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

The input can also be constant marginal return, as in Hanushek & Zhang (2009) , $f(\cdot)$in linear form:

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

The second is that the quality of education is a kind of output, and the accumulation of outputs equals the total output. In the context of this article, this means that the quality of education is the cognitive skill score that can be obtained each year. Therefore, the cognitive skill production function can be directly simplified to [47]:

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

By selecting appropriate function forms and imposing specific assumptions from different perspectives, the quality of education can be obtained.

After estimating the quality of education using the cognitive skills production function, the quality adjustment coefficient can be constructed using the formula to calculate the adjusted average years of education.

5.2.3. Calculate QAYS: The general steps are the divergence point

Therefore, the general sequence of this type of research is to use student

[46] This can also be understood as estimating the average quality of the education system.

[47] In this case, if the quality of education can vary across grades, the production function can be written in additive form:

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

cognitive skills data and formula (26) or (27) to estimate the education quality q of each country in the corresponding period; then, based on the obtained education quality of each country, use formula (29) to calculate the education quality adjustment coefficient $Quality^{benchmark}$; finally, use formula (28) to obtain QAYS data using the average years of education data of each country in the corresponding period.

In the calculation process, we used the number of students' cognitive skills at a certain point in time, such as the scores of 15-year-old students in PISA 2018, and estimated the quality of education from the students' cognitive skills data [48]. As well as the average years of education data at a certain point in time, the time point of adult data varies. Some use the data of the newly graduated population close to the student time, such as the average years of education of the 25-29-year-old population in 2018 (Filmer et al., 2020), and some use the average years of education data of the 25-64-year-old population in the corresponding period (Kaarsen , 2014 ; Altinok & Diebolt , 2023). The choice of the population group with average years of education for adults determines whether QAYS measures a country's current education system or the country's human capital stock. But no matter what choice is made, the student data and adult data are not of the same period. However, theoretically speaking, the data of adults aged 25-29 is closer to the student data in time and is more convincing. Therefore, the most suitable use of QAYS is to measure the quality of the current education system.

In other words, the premise of calculating QAYS is to have data at two points in time: data on students' cognitive skills at one point in time and data on the average years of education received by adults at another point in time. Students' cognitive skills are used to estimate the quality of education, and then the average years of education received by adults are adjusted.

From the above deduction process, it can be seen that the QAYS index is derived after imposing layers of assumptions on the combination of education quantity and quality and the form of human capital production function. This means that the implementation of the research depends on the assumption of cognitive skills production function. At present, the academic community has insufficient knowledge

[48] Ideally, we should use the same point in time to combine the quantity and quality of education. In theory, we should use the amount of education at age 15, but this is redundant, after all, students' cognitive skills are already an indicator that includes both the quantity and quality of education.

of the technology and methods of human cognitive skills production, which makes scholars have to make many subjective assumptions that are not supported by much empirical evidence when conducting relevant research. Therefore, there is a great risk of measurement bias in the construction of the QAYS index. The "black box" of cognitive skills production function is an insurmountable mountain in front of scholars in this field .

VI. Adult cognitive skills : the best measure of human capital

This article discusses the average years of education, students' cognitive skills, and the combination of education quantity and quality. This raises a question: Are the above measurements the best indicators of human capital? If not, which one would be the best indicator?

This paper argues that among all indicators, the best measure of human capital is adult cognitive skills.

There are several reasons for this: First, students' cognitive skills are flow indicators, while adult cognitive skills data are human capital stock indicators; on the one hand, this is because participants in adult cognitive skills tests cover all age groups in the labor market, while students' cognitive skills are usually only for a certain grade or age; on the other hand, students' cognitive skills data are for a certain grade or age, not their cognitive skills after they have completed all their education.

Second, according to the education production function, both the quantity of education (average years of education) and the quality of education are inputs, and the cognitive skills of adults are the final outputs. Third, human capital is the knowledge, skills and other elements condensed in people. Therefore, the human capital measurement index (QAYS) synthesized by the quantity and quality of education is obviously not as good as the direct measurement of human capital.

6.1. Building an adult skills database: Using links with student data to overcome the limitations of adult surveys covering a limited number of countries

Existing adult cognitive skills surveys [49]include the International Adult Literacy

[49] For brief descriptions of more extensive investigations of adult cognitive skills, see De La Fuente & Doménech (2024) and Reiter et al. (2020) .

Survey (IALS) [50], the International Adult Literacy and Life Skills Survey (ALLS) , the [51]Programme for the International Assessment of Adult Competencies (PIAAC) , [52]and the Skills Towards Employability and Productivity (STEP) [53]program . PIAAC covers the largest number of countries, but only provides data for 36 countries .

It can be said that the limited number of countries participating in PIAAC has greatly restricted its use in macro research. Correspondingly, with the development of student survey and testing projects, more and more countries are included. For this reason, a few scholars have recently begun to try to link the cognitive skills of students and adult cognitive skills data in various countries, and expand the number of national samples of adult cognitive skills data by building a conversion relationship between the two (Égert et al. , 2024).

6.2. Birth cohort matching and correlation function for students and adults

In the research direction of expanding adult cognitive skills data, due to the limitation of vision, the author has only found one study: Égert et al. (2024) . Therefore, this section mainly introduces the methodological ideas of this study.

The idea of this method is similar to the relationship between flow and stock. After receiving education, students, as flow, continue to enter the labor market and become part of the adult population; while adults, as stock, are constantly replaced by flow as they continue to integrate into the flow. Therefore, today's adult labor force stock can be seen as the result of continuous replacement by students from previous periods. Therefore, if you have flow data for all periods that replace all adults (that is, the cognitive skills of students in all previous historical periods), you can use the past flow to estimate today's stock (that is, the current cognitive skills of adults).

shown in the figure, suppose that c 15-year-old students in a certain country

[50] The survey was conducted by the OECD between 1994 and 1998, covering 22 countries. The survey content included prose literacy, document literacy, and quantitative literacy.

[51] The survey was conducted by the OECD between 2003 and 2007 and was regarded as the successor of IALS. A total of 11 countries were surveyed. In terms of content, numeracy replaced the previous quantitative literacy, and problem -solving was added .

[52] The survey was conducted by the OECD between 2011 and 2018, covering a total of 37 countries. In terms of content, it consists of literary literacy , numeracy , and problem - solving in technology - rich environments .

[53] The survey was conducted by the World Bank in 17 non-OECD low- and middle-income countries between 2012 and 2017. Unlike the survey conducted by the OECD, the survey content mainly focused on reading skill.

participated in the PISA test from 1995 to 2015, and the test score is flow, and the adults in the country participated in the PIAAC test in 2017, and the test score is stock. If the cognitive skills of 15-year-old students who participated in the PISA test from 1995 to 2015 will not change after participating in the test, we can construct the cognitive skills data of 15-39-year-old adults in the country in 2017 based on the birth cohorts corresponding to the 15-year-old students in the PISA test from 1995 to 2015 in $PISA'$2017. In other words, the cognitive skills of adults aged 15-19 in 2017 in the figure $PISA'$are equal to the cognitive skills of 15-year-old students who participated in PISA 2015, and the cognitive skills of adults aged 20-24 are equal to the cognitive skills of 15-year-old students who participated in PISA 2015, and so on (as shown by the solid arrows in the figure).

the 2017 we constructed above $PISA'$will not really be equal to the adult cognitive skills of Country C in 2017. In order to solve this problem, we need to know the relationship between the constructed adult cognitive skills ($PISA'$2017) and the real cognitive skills (PIAAC2017). Through the corresponding data of each birth cohort in the constructed $PISA'$2017 and the real PIAAC2017 (as shown by the dotted line in the figure), we can estimate the connection conversion equation. Based on this equation, we can use the PISA scores of students' cognitive skills in previous periods to infer the current adult cognitive skills PIAAC.

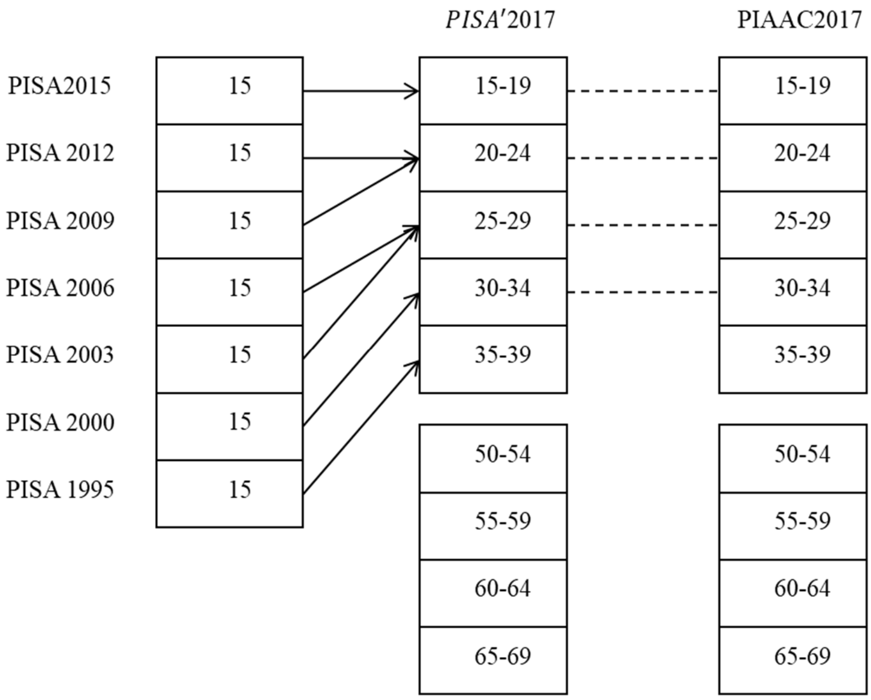


Figure 6-1 Simple schematic diagram of research ideas

According to Égert et al. (2024) , the above operation process is generally divided into three steps :

The first step is to select countries that participate in both PISA and PIAAC as anchor countries . According to the time when students took the test, we can infer/fabricate the adult cognitive skills data for the corresponding period, such as $PISA'$2017 in the figure above.。

The second step is to use the fictitious adult cognitive skills data of the above countries (such as $PISA'$ 2017) and their real adult cognitive skills (such as PIAAC2017), with birth cohorts as data units, to construct a conversion function. This is the most important step, and the quality of the conversion function is directly related to the quality of the subsequent conversion data.

The third step is to apply the constructed conversion function to analyze the data of countries that have participated in many PISA tests, and estimate the cognitive skill scores of adults in each birth cohort in these countries. Then, by performing the conversion on all birth cohort scores in the country, we can obtain the mean or other statistics of the country's overall adult cognitive skills. Since the estimated adult cognitive skills data is for 2020, the countries that participated in the PIAAC test are also included in the estimation.

Égert et al. (2024) pioneered the use of historical human capital flows every year to cumulatively calculate the current adult human capital stock. However, this method also has its limitations. It is more suitable for high-income economies with high basic education enrollment and promotion rates. For countries and regions with underdeveloped economies and education, because a large number of people have not received basic education, the PISA scores of students in these countries cannot represent all the populations of their corresponding birth cohorts. Therefore, the use of this method may lead to an overestimation of the cognitive skills of adults in developing countries.

Using the above method, Égert et al. (2024) constructed a human capital stock database for the population aged 15-64 in 17 countries and the population aged 15-39 in 54 countries in 2020. The number of countries covered by the population data for the

15-39 age group is 18 more than that of PIAAC. In practice, the implementation of this method depends heavily on whether the basic education students in the sample countries have received international tests in the past and how many rounds of international tests they have received. High-income economies are "regulars" in international student tests. Therefore, the sample data constructed by Égert et al. (2024) is still dominated by high-income economies, with only a small number of middle-income economies. Among the countries covered by the population data for the 15-39 age group, high-income economies account for 74.07% and OECD countries account for 68.52%.

VII. Labor Market Wage Information: A Direct Reflection of Education Quality and Human Capital

7.1. Research Intention: Separating Education Quality and Human Capital Information from Wage Information

In theory, the quality of education will be directly reflected in the performance of graduates, and the wages of labor in the market can be used as a direct reflection of personal human capital (Lee & Barro, 2001). Therefore, the wage information of labor in the market can be used to reflect the quality of education and human capital of each country. There are currently two ways to obtain these two types of information from wage information: one is based on the Minser income equation, using the rate of return on education to measure the quality of education in each country. The other is based on the macro production function, separating the impact of human capital from wage income.

Table 7-1 Two ways to use wage information

Base	Separated measurement bias
Based on the Minser income equation	Education Quality
Based on the macro production function	Human Capital

One problem with these two methods is that the wage performance of labor in different countries depends not only on the school education and human capital of each country, but also on the external environment (Lee & Barro, 2001); and the inconsistency of the labor market environment in different countries makes the wage information between countries not comparable. The gap in labor income does not only reflect the difference in labor human capital, but may also reflect the differences in the labor market, such as total factor productivity, the level of job information available, etc. Therefore, existing research usually uses data on immigrants from other countries in the labor market of a certain country in sample selection, which ensures the consistency of the labor market.

Two issues must be considered when using immigration data: First, immigration may be selective, so immigrants from a certain country cannot represent the overall situation of that country. This selectivity mainly comes from two aspects. One is the self-selection of immigrants. People with higher education from low-income countries are likely to immigrate to countries with richer economic development levels. Therefore, using samples with higher personal abilities may overestimate the quality of education in that country; the second is the choice of the country of immigration. For example, countries such as the United States and the United Kingdom usually set some conditions to screen immigrants. Secondly, the question of whether skills can be fully transferred, that is, whether the skills learned in the home country become no longer applicable after immigration. If this is the case, the estimated market rate of return cannot represent the true quality of the country's education system.

7.2. Approach: The Minser Income Equation under the Micro Path and the Human Capital Price under the Macro Path

7.2.1 Mingser Income Equation

When using a country's immigration data to construct the Minser income equation, the following formula is obtained:

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

Where $c2$refers to the country of immigration, $c1$refers to the country of origin of immigrants; $log(W)$is the logarithm of personal income, $S$is the number of years of education of an individual, μrepresents the estimated coefficient of interest, which represents the quality of education in the country of origin of immigrants, and represents the income increase that can be brought about by one year of education; $X$are other control variables.

Using data on immigration within a country (such as the United States), it is possible to estimate the quality of education in each country μ^c.

Currently, there is limited research using this method to separate educational quality. Regarding the two issues of immigrant selection and skill transferability in immigration data, existing research has only verified that immigrant selection and skill transferability issues will not have a significant impact on the estimated results (Schoellman , 2012).

7.2.2 Human capital price

This method starts from the macro production function. The standard production function is:

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

Among them Y_c, is total output; $K_c^{\alpha}$is the stock of physical capital; $A_c$is total factor productivity; $H_c = h_c L_c$is total labor input, $h_c$is per capita human capital, $L_c$and is the number of workers.

For production, in order to maximize profits, the following formula needs to be maximized:

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

The first-order derivative shows that the wage per unit of human capital (i.e., the price of skills) is:

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

It can be seen that the skill price (w_c) of each country is affected by the total factor productivity and capital output ratio of each country. In addition, if workers are paid wages based on marginal output, then the workers' income is:

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

thus,

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

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

Among them $\log(h_c) = E_c[\log(h_{i,c})]$. From the above formula, we can see that if we know the human capital of each country z_c, we can obtain the human capital of each country through fixed effects.

Therefore, this type of research is usually divided into two steps. The first step is to estimate z_c; the second step is to use the above formula to obtain the human capital of each country.

Let's take the simplest case as an example: when there is no problem of immigrant selection and skill transferability, the human capital of immigrants does not change with immigration, so:

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

By estimating the above formula, we can obtain the for each country z_c. However, this estimation has extremely high data requirements, requiring wage data before and after immigration and multiple source countries and destination countries of immigrants [54].

The problem becomes complicated when there are issues of immigration selection and skill transferability. Some studies have addressed this issue by internalizing immigration selection and skill transferability (Martellini et al. , 2024).

Ⅷ. Human capital indicator analysis framework

We have introduced the evolution history of human capital indicators and the five key measurement indicators in the evolution process. However, the existing analysis of the indicators is rather scattered and not put into the same framework.

8.1. Human capital indicator analysis framework in a unified framework

8.1.1 Theoretical framework

According to human capital theory, human capital refers to the sum of quality factors such as knowledge, skills, and physical strength (health status) that have

[54] It should be noted that when all immigrants come from one country, or when only immigration data for one country are obtained, this formula can only estimate the difference between the skill prices of each country and the immigration country, but this difference does not affect the human capital estimates of each country.

economic value and exist in the human body .

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

However, the existing measurement of human capital is more about the measurement of skills, so the above formula is usually:

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

According to the usual classification, skills are divided into cognitive skills and non-cognitive skills:

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

Due to the limitations of measurement technology, most existing methods focus on cognitive skills, so there are:

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

The production of skills, especially cognitive skills, can be divided into two stages: before graduation and after graduation.

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

There $afterschool$may be a huge time gap between them. After graduation, skills are mainly student skills ($Skills_{atschool}$) $atschool$, influenced by on-the-job education ($onjob\ education$), experience ($experience$) and depreciation ($aging$).

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

Before graduation, according to the education production function, cognitive and non-cognitive skills are outputs, which are affected S by factors such as school education ( ), family input (F), and talent ( ). According to AHanushek & Woessmann (2012a) and Hanushek et al. (2015) , the formula can be written as:

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

The division between before and after graduation serves two purposes. First, the influences before and after are not the same; second, it is to emphasize the long distance in time and prevent the confusion of skills at different stages at the same time point: the cognitive skills of an adult in 2025 are influenced by the education they received as a student, not by the education in 2025. The above formula $S$is the influence of school, which can be seen as a function of the quantity and quality of school education:

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

From another perspective, the quantity and quality of school education are also the output of the input of various schools:

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

Among them, $ExpendPS$represents the per capita funding, $Teacher$represents the teaching staff, and X represents other inputs.

8.1.2 Indicator role

Based on the above framework, we can put direct-indirect, stock-flow, output-input, quality-quantity, and labor market-education system into this framework.

(Direct-Indirect) First, direct and indirect. If there is a so-called optimal measurement of human capital, it must be directly related to people' s knowledge and skills. Any measurement indicator that does not involve factors such as knowledge and skills, health, etc. is only an indirect measurement of human capital.

(Flow-Stock/Labor Market-Education System) Secondly, when we talk about a country' s human capital, we usually refer to the human capital of all adults in the labor market. Therefore, to measure a country' s human capital, we should measure the stock of a country' s human capital. For groups that have not yet entered the labor market, such as students in the education system, their source of human capital stock is a flow indicator that has not yet occurred.

(Output-input) Cognitive skills are an important component of skills. The adults in a country's labor market at a certain time are composed of people from different birth cohorts. According to the education production function, **the cognitive skills of a population of a certain birth cohort at a certain time can be regarded as a function of the cognitive skills acquired by adults of that age group during their school years; (quantity-quality) The cognitive skills acquired during the school years can be regarded as a function of the minimum number of years required to complete all levels and types of education stipulated by the country at that time (i.e., the quantity of education) and the quality of education received (the quality of education can be regarded as the average quality of the education received); and**

the quantity and quality of education are the product of a country's long-term development of education, the result of the country's long-term investment in education, and a function of relevant education inputs at all levels and types (such as teacher-student ratio, per capita funding). Under different circumstances, the same factors (quantity and quality of education) can be both inputs and outputs. **If we regard educational production as a dynamic process in which educational inputs and outputs are continuously transformed at different levels, then in the final stage, the final product of a country's human capital investment should be the cognitive skills of the adult population, with the cognitive skills acquired during student years being the input; in the middle stage, the cognitive skills acquired during student years are the output, and the number of years and quality of education received by students are the input; in the most basic stage, both the quality and quantity of education are outputs, while per capita funding, enrollment rates, and student-teacher ratios are inputs.**

8.2. Classification and characteristics of human capital measurement indicators

Analyze the human capital measurement indicators under the framework. Before forming the framework, it is necessary to summarize the indicators that appear.

8.2.1 Output-input, direct-indirect, stock-flow, quality-quantity

Based on the keywords described above, we summarize the mentioned indicators by direct-indirect, stock-flow, input-output and quality-quantity.

Table 8-1 Human capital measurement indicators under direct - indirect and stock - flow classifications

	Stock indicators	Traffic indicators
direct	Adult cognitive skills (mean)	Student cognitive skills (mean)
indirect	QAYS (all population) Separating human capital from wage information Average years of education Literacy rate	QAYS (a certain birth cohort) Disentangling educational quality from wage information Teacher-student ratio, per capita funding, enrollment rate, etc.

Table 8- 2 Human capital measurement indicators under the input-output and quality-quantity divisions

level	Input OR	Quantity indicators	Quality indicators

<table>
<tr><td></td><td>Output</td><td colspan="2"></td></tr>
<tr><td rowspan="2">Top floor</td><td>output</td><td colspan="2">Adult cognitive skills (mean)</td></tr>
<tr><td>Investment</td><td colspan="2" rowspan="2">Cognitive skills acquired by adults during school years, QAYS (a particular birth cohort)</td></tr>
<tr><td rowspan="2">Middle level</td><td>output</td></tr>
<tr><td>Investment</td><td rowspan="2">Education quality, student cognitive skills (mean) (grade or grade-consistent) separated from wage information</td><td rowspan="2">Average years of education</td></tr>
<tr><td rowspan="2">Bottom Layer</td><td>output</td></tr>
<tr><td>Investment</td><td>Teacher-student ratio, per-student funding</td><td>Enrollment Rate</td></tr>
</table>

8.2.2 Labour market and education system

Although the above indicators can all be used to measure a country's human capital, some indicators tend to measure a country's education system indicators. In the wording of the article, we try to distinguish between human capital and education quality. In the table below, we summarize the tendencies of these indicators. The important and necessary reason for this distinction comes from the attention to the time dimension information. The impact of the education system on the overall human capital is gradual, slow and lagging. Therefore, if a correspondence is to be made, the time of the education system indicators should be earlier than the human capital measurement indicators, and some existing research treatments ignore this point.

We believe that the average years of education is both a measure of human capital and an indicator of the education system, because education is used purely to measure human capital; student cognitive skills are indicators of the education system, because the objects of their measurement are mostly students in primary and secondary education; the adjusted average years of education, depending on whether the average years of education are for the entire population or the birth cohort population that has just entered the labor market, are biased towards human capital and the education system, respectively, but it should be noted that, in theory, the quality of education used for adjustment is obtained from the student cognitive skills data, because this method theory should be biased towards the education system; adult cognitive skills are the cognitive skills of the population on the labor market, so as advocated in this article, they are the optimal proxy indicators of human capital; wage information on the labor market is used to separate out the quality of education, so it is also biased towards the education system.

Figure 8-1 Schematic diagram of the tendency of various measurement indicators under the human capital - education system division

8.3. Human capital measurement indicators based on a unified framework

8.3.1 Principles for evaluating the quality of human capital measurement indicators

of human capital indicators in the previous article already reflects the tendency of evaluating human capital indicators. This article summarizes these tendencies into five evaluation principles: "comprehensive measurement is better than one-dimensional 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".

First, human capital has a rich connotation, including skills, knowledge, health, etc. Existing measurements of human capital mostly focus on skills, especially cognitive skills. Skills or cognitive skills are only one dimension of human capital, not all of it. The optimal measurement should include all the connotations of human capital.

Second, if there is a so-called optimal measurement of human capital, it must be directly related to people's knowledge and skills. Any measurement indicators that do not involve knowledge and skills are only indirect measurements of human capital, which means that "direct indicators are better than indirect indicators." From this perspective, if we want to measure a country's educational human capital, we only need to measure the cognitive skills of adults in that country; if we want to investigate the quality of a country's education system, we only need to measure the value-added changes in the cognitive skills of school-age children before and after receiving formal education.

Third, human capital can promote economic growth. Although it is not explicitly stated, the human capital here should refer to the human capital of all people. Therefore, in theory, all adults in the labor market should be measured directly (i.e., stock measurement) rather than students (i.e., flow measurement). The impact of a country's educational development on the human capital of its population is lagging. Existing studies have also confirmed that the distribution of cognitive skills of basic education students and adult cognitive skills in most countries in the world usually changes significantly. Therefore, it is questionable to use the cognitive skills of current students to represent the cognitive skills of adults in the same period (Huang Bin et al., 2024). Therefore, "stock indicators are better than flow indicators."

Fourth, in **the dynamic process of transforming education input-output at different levels** , although various input and output indicators have been used to measure a country's educational human capital, we should know that the measurement of output should be higher than the measurement of input, because it is inappropriate to regard all inputs as effective productive investments (Hanushek , 2003), and there is uncertainty about how much output the input can produce (Schoellman, 2012; Hanushek, 2003) . In this case, it is more correct to measure the output directly, so "output indicators are better than input indicators."

Fifth, as human capital theory has been widely disseminated and accepted, almost every country in the world has developed education and human capital. The development of education should focus on both the quality and quantity of education. Improving the quality and quantity of education is to improve the final cognitive skill output. If a country provides extremely low-quality education to its people, even if the average years of education for its citizens are more than 16 years, the cognitive skill level of its citizens will not be too high. Such education investment is inefficient and will not generate economic value for external social and economic development. Similarly, if a country provides a limited amount of education to its people, even if the quality of education is excellent, the cognitive skill level of its citizens will not be too high, because the impact of one year of education cannot be too great, and the economic value of such education investment to the external economy and society will eventually be limited. Therefore, for the long-term development of a country, expanding the number of education and improving the quality are equally important. However, as the general concept of investing in human capital spreads around the world, a lot of information has changed and distorted, losing a lot of essential content and power

(Hanushek & Woessmann , 2015). Policymakers and scholars have not really paid attention to the essence of human capital: knowledge, skills, etc., but have focused on proxy indicators related to education level, such as average years of education, enrollment rate and other quantitative indicators. The reality that quantitative indicators such as average years of education and enrollment rate are used as the focus of government policies tells us that some countries (such as Latin American countries) have expanded educational opportunities and increased average years of education with the help of themselves and other countries, but economic growth is still slow, and there is no obvious sign of catching up with developed countries (Hanushek & Woessmann , 2008). These indicators mask the differences in education quality among countries; and "schooling is not learning" (Pritchett, 2013 ; World Bank, 2018; Kaffenberger & Pritchett, 2017; Filmer et al., 2020; Angrist et al . , 2021). Only by transforming schooling into effective learning and improvement of students' real skills can education development be productive, and thus have a sustained endogenous driving effect on a country's long-term economic growth (Huang Bin and Yun Ruxian, 2023). Therefore, in this context, "quality indicators are better than quantity indicators".

8.3.2 Analysis of various human capital indicators

From the perspective of direct and indirect measurement, (student/adult) cognitive skills indicators are better than QAYS or LAYS, years of education and other indicators; from the perspective of stock and flow, adult cognitive skills indicators and years of education are better than student cognitive skills and enrollment rate indicators; from the perspective of output and input, adult cognitive skills indicators and other indicators are better than teacher-student ratio, per capita funding and other indicators; from the perspective of quality and quantity, student cognitive skills indicators are better than years of education, enrollment rate and other indicators. Based on the above comparisons, adult cognitive skills should be the best indicator for measuring a country's educational human capital, and it has an absolute advantage over other indicators.

Secondly, we believe that if the goal is to measure human capital, labor market indicators should be given priority. If direct measurement is available, direct measurement should be given priority. If there is no direct measurement, stock measurement should be given priority. If there is no stock measurement, output indicators should also be given priority. Among input indicators, quality measurement should be given priority.

(1) Rediscussion of the relationship between adult cognitive skills, years of education, and student cognitive skills

It is necessary to reiterate the understanding of the relationship between adult cognitive skills, average years of education, and student cognitive skills. These three are the focus of existing research, and increasing the understanding of them will help understand existing research.

According to the analytical framework, the following formula can be obtained:

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

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

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

Adult cognitive skills are a function of the cognitive skills of adults in their school years. It should be noted that there is a huge difference between cognitive skills in school years and student cognitive skills. First, the two are not consistent in time. The school years of today's adults are many years ago, while today's student cognitive skills are today's. Second, in concept, the cognitive skills in their school years are the final result of their education. This is ideal, but in fact there is no measurement. If it is measured, it should be measured when they finally leave school. Because different people leave school at different times, this makes it difficult to implement. The actual measured student cognitive skills, due to measurement reasons, refer more to students of a certain age or grade who are still receiving education, and their cognitive skills are not the final result.

Due to the above distinction, strictly speaking, the following formula does not hold:

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

Secondly, as for students' cognitive skills, they can be regarded as the quality of education as long as the quantity of education is consistent, that is, the assessment is conducted on students of the same age or grade. Otherwise, students' cognitive skills are not equal to the quality of education.

IX. Summary, Evaluation and Outlook

Reliable and accurate measurement of human capital in various countries is the basic work for conducting macro-education policy research. This article systematically introduces the human capital measurement indicators such as education quality and human capital separated from years of education, students' cognitive skills, QAYS, adult cognitive skills and labor market wage information, along the "evolutionary history" of human capital measurement, including their existing problems, construction principles and methods, and commonly used international databases. After systematically introducing the mainstream indicators of national education human capital, it is necessary to conduct a more in-depth summary and comparative study of these four indicators, analyze their respective problems, condense the basic principles for judging the pros and cons of various measurement indicators , and point out the development direction of future research.

9.1 Summary: The development law of human capital measurement indicators

From the development process of human capital measurement indicators, we can clearly find the following rules: First, output indicators are preferred to measure education or human capital, rather than input indicators. Average years of education and cognitive skills are direct output indicators. Measuring the education system or human capital from the output side rather than the input side is a result-oriented result, which is the result of thinking about "Can input bring output? How much output can it bring?"

Second, the emphasis has shifted from quantitative indicators to more and more qualitative indicators. The shift in human capital measurement indicators from average years of education to cognitive skills fully reflects this. This is due to the fact that the existing policies aimed at quantitative indicators such as average years of education have not brought about the expected economic growth, as well as the differences in education quality between education systems and people's understanding that schooling is not the same as learning .

Third, in macro research, researchers have changed from only considering quantity to considering both quantity and quality of education. However, it should be noted that for individuals, this means receiving both a sufficient amount of education and a high enough quality of education; for a country, this means "a large number of educated people but insufficient education quality" and "excellent education quality but limited educated people" will both limit the development of a country's productivity.

Fourth, although in recent years, too much attention has been paid to flow indicators (student cognitive skills) in cognitive skills data, there is a trend of returning to stock indicators (adult cognitive skills). Human capital stock generally refers to the human capital of all labor forces. Compared with human capital flow, it has a higher causal relationship with relevant indicators such as the economy (De La Fuente & Doménech , 2024).

9.2 Evaluation: Human capital measurement indicators and data have limitations

In the analysis of Chapter 8, we assume that all human capital indicators are perfect and there are no problems with their measurement. However, from Chapters 3 to 7, we can also see that due to data construction,

9.2.1 Years of education

In order to obtain comparable average years of education between countries, it is necessary to unify and estimate the data when constructing it, such as unifying the different education stages of each country into several broad stages. Since the statistical data published by each country is based on its own international norms, the above operation will cause the constructed data results to be different from the statistical data published by each country, so it is not appropriate to directly compare the constructed data with the data published by each country. In addition, the biggest flaw is just as it has been criticized, it cannot reflect the differences in the quality of education received.

9.2.2 Cognitive skills

Using cognitive skills to measure education quality or human capital has become an important current research and development direction. However, it is undeniable that there are still many limitations in terms of current data.

In terms of measurement content, there is currently no test that can cover and measure all the abilities and skills that determine a country's innovation capacity and labor productivity, including non-cognitive skills, skills acquired in universities and workplaces, and highly specialized and complex knowledge and skills of scientists and high-level technical personnel (De La Fuente & Doménech , 20-24). In terms of measurement objects, existing international cognitive skills test projects are more about measuring students' cognitive skills information, so they can only be used to measure the quality of primary or secondary education. There is currently no cognitive skills test project data to measure the quality of higher education; for countries with low enrollment rates, these student cognitive skills test project scores may only represent the knowledge and skills of some people in the birth cohort participating in the test project, rather than the knowledge and skills of the entire birth cohort; in addition, students have not yet entered the labor market, so it is still insufficient to use them as a substitute for the labor force as an indicator of the education quality or human capital of the entire country. In theory, there are serious problems of endogeneity and mutual causality (Huang Bin et al., 2024 ; De La Fuente & Doménech , 2024) [55]. In terms of measurement countries and time, we are far from having enough information to measure human capital (and changes) over long periods of time in most countries around the world. Even for student testing projects, many countries have only participated in the most recent international testing projects, so only a few countries have long-term series data. In addition, the quality of existing cognitive skills data needs to be further improved. Although all current studies have mentioned the risks that may exist when the content identity assumption is not strictly met, no research has considered how to solve the problem of test content differences through certain methods when the content identity assumption is not strictly met, so as to obtain more consistent conversion data and improve the quality of conversion data.

When evaluating students' cognitive skills as educational acquisition, scholars usually only consider the comparability of cognitive skill scores and ignore students' uniformity (such as age).

9.2.3 Improve the average years of education

The reform of the average years of education includes both quantity and quality dimensions in the same indicator. In this way, the commonly used average years of education is not abandoned, but the quality dimension indicator of cognitive skills is also added.

[55] In this case, studies using the average years of schooling of the labor force may be less susceptible to reverse causality problems (De La Fuente & Doménech , 20 24). Of course, the most direct way is to directly use adult cognitive skills (Huang Bin et al., 2024).

However, this indicator contains a large construction error. The implementation of the study relies on the assumption of the cognitive skills production function (it is undeniable that this study has deepened people's understanding of the cognitive skills production function), and the lack of understanding of cognitive skills production will limit the accuracy of the final results obtained by this method; to overcome the lack of understanding of the cognitive skills production function, it is necessary to measure the cognitive skills of each age group and observe the general law of cognitive skills growth. This should take into account the influence of education as well as physiological and psychological maturity. For these, researchers need to make more efforts. In addition, when using adult cognitive skills to estimate the quality of education, due to the depreciation of cognitive skills, as well as the influence of factors such as on-the-job training and learning by doing (De La Fuente & Doménech , 2024), it makes it more difficult to obtain a clean education system quality, and the error is greater when estimating. In addition, it is easy to overlook that there are also measurement errors in the measurement of cognitive skills, which will accumulate to the average years of education after the transformation .

Secondly, the role of this indicator is limited. When the quality of education is consistent at each stage of education, if you want to measure the quality of education, the cognitive skill score of students in a certain grade is a better indicator; when the quality of education is inconsistent at each stage of education, if you want to measure the quality of education, the value-added change of students' cognitive skill scores is a better indicator; when measuring the quantity of education, the average years of education is a better indicator; when measuring the final educational output, the final adult cognitive skill score is the best indicator. Under this framework, there is no need to reform the average years of education. In order to better serve policy goals, education statistics and surveys should be strengthened to accurately grasp the data on the years of education of the population and the cognitive skills of the population of all ages.

9.2.4 Salary information on the market

Although it is feasible to separate education quality or human capital from wage information on the labor market when immigration data are available, this method is at a disadvantage when compared with average years of education and cognitive skills data due to its own particularity, which limits its application in measuring education quality or human capital.

First, in terms of considerations, the use of wage information in the labor market requires the study to consider some factors after students enter the labor market, which makes this type of research more difficult to implement than previous studies in theory.

Secondly, in terms of data usage in practice, it has higher requirements for data. On the one hand, this study needs to use immigration data, which is difficult to obtain; and its results are easily affected by data from different sources. Although the results obtained from immigration data in different countries are highly correlated, it cannot be denied that the results are unstable. On the other hand, the quality of the education system it measures is the weighted value of individuals entering the labor market. Since this type of research assumes that the quality of the education system remains unchanged, it ignores the consideration of time. Therefore, ideally, it should include enough representative samples of each period. If the individual samples of some countries observed entered the labor market early, their market return rate may measure the quality of the country's education system in the early days, which may actually be incomparable with other countries.

Finally, in terms of temporal changes, the indicators obtained by this type of research cannot measure the changes in education quality and human capital over time. This type of research is based on the income data of immigrants with different graduation times and similar working hours, and obtains an estimate of a country's education quality or human capital that does not change over time. However, the quality of education in a country is different at different times, which makes the individual return rate of receiving the same amount of education and the human capital obtained by receiving education at different times different. At the same time, this method makes it difficult to compare the differences in the quality of the education system or human capital over time.

Due to the above limitations, the educational quality or human capital of this type of research is difficult to use in policy practice.

9.3 Outlook: Suggestions on the use of human capital measurement indicators

First, distribution always contains more information than a single mean. This is true for both educational achievement and cognitive skills. In educational achievement, the distribution of educational attainment (the proportion of the population that reaches a certain level of education) contains more information than the single average years of education; in cognitive skills, the distribution of cognitive skills (mean, quantile value, standard deviation, skewness) contains more information than the single mean of cognitive skills. Although Hanushek & Woessmann (2012a) , Huang Bin and Yun Ruxian (2023), and Huang Bin (2024) have attached great importance to distribution, (2024) has made some progress, but it still needs to be explored in depth. The use of distribution information is not only a full utilization of existing information, but also an innovative development of human capital theory in empirical research, which means that the research extends from the simple overall level of human capital to the richer connotations of human capital difference characteristics and structural characteristics. In addition, in the construction of a comparable database of students' cognitive skills, the applicability and accuracy of various methods in the conversion of other distribution characteristics need further analysis and discussion.

Second, the changes in education quality over time should not be ignored. In the context of rapid expansion of education, especially higher education , it is particularly important to study the changes in human capital over time. The basic assumption of most studies is that the quality of education remains unchanged (De La Fuente & Doménech , 2024). However, during the period of rapid expansion of education, the quantity and quality of education will change dramatically. At this time, if the time dimension is not considered in the measurement of human capital, the reliability of the research results will be affected; simple empirical analysis also shows that in many countries, the quality of education has not remained stable for a long time (Hanushek & Zhang , 2009). If we want to realize the time change of a country's education human capital stock, we cannot only focus on the mean level of population cognitive skills in a single historical period in each country. We must group and statistically describe the cognitive skills of different birth cohorts in different historical periods, and fully understand the historical evolution of major reforms in the education mechanism and system of each country, and estimate the causal effect of these reforms on the changes in the stock of education human capital in different historical periods.

Third, although adult cognitive skills are the best measurement indicator of a country's human capital, the use of cognitive skills to measure education quality or human capital needs to be further strengthened. Whether it is student cognitive skills or adult cognitive skills, they are not just the product of school education. So far, no research has been able to give an accurate answer to the impact of extracurricular education on cognitive skills. In addition, although research often uses student cognitive skills as a measure of the quality of education in various countries, the production of student cognitive skills should also be the joint output of both the quantity and quality of education. Therefore, the cognitive skills of students of the same grade (the same number of years of education) should be used, and data from multiple grades should not be mixed. Although recognizing this point does not affect previous research, it is not often emphasized.

Supplementary data

In the appendix, we provide a literature summary of each method, commonly used database results, and a detailed introduction to the methods in some literature.

References

[1] Huang Bin, Yun Ruxian . (2023). How can educational development strengthen a country? An empirical analysis based on internationally comparable data on cognitive skills from 1960 to 2020. Educational Research, 4 4(10), 125-136 .

[2] Huang Bin, Yun Ruxian , Wu Kailin. (2024). The impact of cognitive skills distribution on national economic growth: New evidence for a strong country through education. Journal of East China Normal University (Educational Science Edition), 42 (9), 13-32 .

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

[4] Altinok , N., & Diebolt, C. (20 2 4). Cliometrics of Learning-Adjusted Years of Schooling: Evidence from a New Dataset. *Cliometrica ,* 18 (3), 691-764.

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

[6] Altinok , N., & Murseli, H. (20 07). International Database on Human Capital Quality. *Economics Letters* , 96(2), 237-244.

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

[8] Angrist, N., Evans, D. , Filmer, D.P. , Glennerster , R., Rogers , FH, & 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 , CA, 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 , RJ , & Lee , J. (1993). International Comparisons of Educational Attainment, *Journal of Monetary Economics* , 3 2(3), 363-394.

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

[12] Barro , RJ , & Lee , J. (20 13). A New Data Set of Educational Attainment in the World, 1950–2010 . *Journal of Development Economics* , 104(September), 184-198.

[13] Barro , RJ , & 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., & KC, S. (2012). Populations for 171 Countries by Age, Sex, and Level of Education around 2010: Harmonized Estimates of the Baseline Data for the Wittgenstein Center Projections. IIASA Interim Report IR-

12-016 .
[15]Braun, HI, & Holland, PW (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In PW Holland & DB 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. (20 10). 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]D e L a Fuente, Á. , & Doménech , R. (200 0). Human Capital in Growth Regressions: How much Difference Does Data Quality Make? *W orking Paper* .
[19]D e L a 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]D e L a Fuente, Á. , & Doménech , R. (20 15). Educational Attainment in the O ECD , 1960–2010. Updated Series and a Comparison with Other Sources . *Economics of Education Review* , 4 8(October), 56-74 .
[21]D e L a 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]E gert , B ., D e L a 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. , KC , 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, EA (2003). The Failure of Input-Based Schooling Policies. *Economic Journal* , 113 (485), 64-98.
[30]Hanushek, E. A. , & Kimko , DD (2000). Schooling, Labor-Force Quality, and the

Growth of Nations . *American Economic Review,* 90 (5), 1184-1208 .
[31]Hanushek, E. A. , & Woessmann , L. (20 08). The Role of Cognitive Skills in Economic Development . *Journal of Economic Literature* , 46(3), 607-668 .
[32]Hanushek, E. A. , & Woessmann , L. (2012 a). 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. (2012 b). Schooling, Educational Achievement, and the Latin American Growth Puzzle . *Journal of Development Economics* , 99(2), 497-512 .
[34]Hanushek, E. A. , & Woessmann , L. (201 5). Universal Basic Skills: What Countries Stand to Gain . Organization for Economic Co-operation and Development, Paris.
[35]Hanushek, E. A. , & Zhang, L. (20 09). Quality-Consistent Estimates of International Schooling and Skill Gradients. *Journal of Human Capital* , 3(2), 107-143 .
[36]Holland, PW, & Thayer, DT (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, MJ (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics,* 9 (1) , 25-44.
[39]Kolen, MJ, & Brennan, R. L. (2014). Test Equating, Scaling, and Linking : Methods and Practices . New York: Springer .
[40]Krueger, AB, & Lindahl, M. (2001). Education for Growth: Why and for Whom? *Journal of Economic Literature* , 39 (4), 1101-1136.
[41]Lee , J. , & Barro , RJ (20 01). Schooling Quality in a Cross–Section of Countries . *Economica* , 68 (272), 465-488.
[42]Lim , S.S. , Updike , R.L., Kaldjian , AS, 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., KC, 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 , HA, & Angrist , N. (2018). Global Dataset on Education Quality: A Review and Update (2000-2017) . *Policy Research working paper* , Washington, DC: World Bank Group.
[48]Pritchett, L. (2013). The Rebirth of Education: Schooling ain't Learning.

Washington, D. C .: CGD Books.
[49]Reardon, SF, Kalogrides , D., & Ho, AD (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. *W orking 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 D cores 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, TM (1961). Investment in Human Capital. *American Economic Review* , 51(1), 1-17.
[55]Singh, A. (2014). Emergence and Evolution of L earning G aps across Countries : Linked P anel E vidence from E thiopia, India , Peru and Vietnam. *CSAE working paper* .
[56]Speringer , M., Goujon, A., KC, S., Potančoková , M., Reiter, C., Jurasszovich , S., & Eder, J. (2019). Global Reconstruction of Educational Attainment, 1950 to 2015: Methodology and Assessment. *VID Working Paper* .