How to measure a Country's human capital?

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

（Online version）

Version: 2025-01-11

Ruxian Yun

（Institute of Education, Nanjing University　210023）

Description:

This is the online English version of the paper "How should a country's educational human capital be measured".

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. The final but essential point is that the online version appendix provides a detailed analysis and summary of the author's previous literature methods (the author wrote the paper based on these 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.

Please cite:

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

This article can currently be obtained from the following website:

Thanks:

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 a Country's human capital?

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

Ruxian Yun, Bin Huang, Yawen Zhu

(Institute 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 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. 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 human capital measurement indicators. measurement indicators, as well as the existing problems, data construction principles and methods, and commonly used databases, along the "evolution history" of human capital measurement indicators, and the "evolutionary history" of human capital measurement indicators. This paper introduces the current commonly used human capital measurement indicators, as well as the existing problems, data construction principles and methods, and commonly used databases, along the "evolution history" of human capital measurement indicators, which mainly include years of education[1] , students' cognitive skills, quality-adjusted years of The human capital measurement indicators, which mainly include years of education , students' cognitive skills, quality-adjusted years of schooling, adult cognitive skills, and

[1] Years of education and students' (adult) cognitive skills are individual micro-statistics, which in another way can be seen as types of data, while average years of education and mean students' (adult) cognitive skills are aggregate macro-statistics, which in another way can be seen as applications of data. However, in macro studies, when referring to students' (adult) cognitive skills, the through point is the students' (adult) cognitive skills mean, which is linguistic convention and can be an obstacle to understanding. For this reason, this paper distinguishes between them as far as possible in terms of terminology. The main reasons for this distinction are, firstly, that micro and macro can be distinguished and linked together; secondly, that the mean is only an aggregate statistic, and that other aggregate statistics such as the GINI coefficient of education, cognitive skill bias, etc., can also be obtained from micro-individual statistics, except for years of education, where the average number of years of education is used the most, and the others to a lesser extent. However, while this paper intends to make this distinction, the reader should also be aware of the meaning in the corresponding context, due to capacity constraints.

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 a number of recommendations. 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 four basic principles for judging the advantages and disadvantages of human capital measurement This paper introduces the classification and characteristics of each measurement, and puts forward four basic principles for judging the advantages and disadvantages of human capital measurement , namely, "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 overall development of human capital measurement indicators. 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

I. Introduction

Human capital is one of the most important factors driving the growth of the national economy, so in order to complete a macroeconometric study or other studies focusing on economic and social development[2] , it is necessary to realize the accurate measurement of a country's human capital first. Although there are many ways to invest in human capital, such as education, migration, health, etc., education is the most important and major way to realize the accumulation of human capital in a country, so in actual research, human capital is often narrowly limited to purely come from investment in education, and therefore, previous studies often use only education-related indicators to measure[3] human capital (Reiter et al., 2020)[4] . To be precise, the resulting measurements point more to educational human capital than to human capital as a whole.

The report on the Twentieth Congress proposes a three-in-one development strategy for education, science and technology, and human resources, fully reflecting the fundamental and strategic role of education in promoting the construction of a strong nation. The strategy for a strong nation emphasizes the productive function of education in external socio-economic development, and calls for education to increase its support and contribution to the realization of the strategic goals of common prosperity and Chinese-style modernization. The development of education has long been regarded as the most important investment for a country to acquire and accumulate human capital, which in turn is an important source of long-term economic growth (Lucas, 1988; Romer, 1986; Schultz, 1961). Thus, the most basic and important question naturally arises: how should a country's human capital be measured?

Measurement issues of human capital are fundamental to macro policy and econometric research, and if we cannot solve the measurement problems related to human capital and ensure its measurement accuracy and validity, there is no way to talk about the validity of subsequent research. It has been pointed out that the inability

---

[2] The main focus is on 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 the education system is often assumed to be measuring human capital, which has led to some confusion in measurement and related terminology, such that measuring the quality of the education system is assumed to be measuring the quality of human capital.

of previous macro studies to reach satisfactory conclusions is, to a large extent, closely related to the large measurement bias of the human capital used (Krueger & Lindahl, 2001; De La Fuente & Doménech, 2006). On the other hand, the academic community has always practiced "fetishism" of human capital measures across countries, with little knowledge of the details of the construction of the human capital measures used , and thus often incorrectly "force" the analysis of data on human capital in education that are not comparable across countries, resulting in a "forced" analysis of human capital in education. As a result, data on human capital in education that are not comparable across countries are often wrongly "forced" to be analyzed, and the results of international comparisons thus obtained are often not credible. In addition, in recent years, especially after 2010, there has been a tremendous development of human capital measurement indicators, with the emergence of indicators such as students' cognitive skills, quality-adjusted years of schooling and adults' cognitive skills , which have been used in a number of studies and have led to conclusions with important policy implications. All of the above makes it urgent and important to review and scrutinize the existing human capital measures, however, to the best of the authors' knowledge, there is no domestic study that focuses on this, and the discussion in the scattered foreign literature is not comprehensive (De La Fuente & Doménech, 2024).

In view of this, this paper firstly comprehends the evolution of national education human capital measurement indicators, and then systematically introduces the commonly used human capital measurement indicators along the development vein, including the data existence problems of various indicators, the principles and methods of data construction, and the commonly used databases. Finally, we summarize the general understanding of these indicators and the overall development pattern, and based on the theory of human capital and the theory of education production function, we summarize and compare and analyze the strengths and weaknesses of the human capital measurement indicators introduced in the paper, and discuss the direction of the development of future national human capital research. It is hoped that with the help of this paper, researchers' understanding of the indicators and data for measuring human capital in education can be strengthened, and readers can be helped to better utilize these relevant indicators and data in research practice.

II. The "evolutionary history" of national human capital measurement indicators.

Since the 1960s, UNESCO has systematically collected data on enrollment rates at all levels of education across countries, and thus early macroeconometric studies often used enrollment rates as a proxy for human capital across countries in studies that empirically analyzed the impact of human capital on aggregate productivity (Hanushek & Kimko, 2000; Barro & Lee, 2013). 2013)[5] . The advantage of the enrollment rate indicator is that it is easy to obtain, but it is obvious that it has a deviation from the theoretical connotation of human capital: human capital is a stock concept, while enrollment rate is an indicator of educational flow, so the response of human capital to enrollment rate is gradual, with a very large lag, and therefore the enrollment rate of a country in a certain period of time does not reflect the level of accumulation of the stock of human capital of the country's adult labor force in that period of time. In contrast, from a theoretical point of view, years of schooling as an indicator of educational output, and thus a country's average years of schooling (Mean Years of Schooling (MYS)) is a much better indicator of the human capital of each country than the enrollment rate, which is a macro-input indicator, and the measurement of a country's adult labor force population's years of schooling and the calculation of the country's average MYS is in fact the measurement of the country's human capital stock is measured. Driven by this idea, international organizations and a large number of researchers have begun to try to construct a database of the average years of schooling of the labor force in each country to achieve a direct measurement of the stock of human capital in each country, and the average years of schooling has gradually replaced the enrollment rate as a commonly used indicator for measuring a country's human capital (De La Fuente & Doménech, 2006).

Average years of schooling as a measure of a country's human capital also has a number of shortcomings, and two criticisms have emerged as the indicator has become more widely used. The first criticism emphasizes that data on average years of schooling from different sources are subject to large measurement errors, which often

[5] Literacy rates have also been used extensively in empirical research in earlier times, but they are more difficult to obtain by comparison. See Berro & Lee (1993) for a detailed discussion of literacy and enrollment rates.

leads to discouraging results in empirical analyses of economic growth using average years of schooling as a measure of human capital (Krueger & Lindahl, 2001; De La Fuente & Doménech, 2006).). A second criticism points to the fact that average years of schooling measures only the quantity of education received by a country's nationals, ignoring the large differences in the quality of education across countries. Macroeconometric studies often use country- and region-level data samples, and it is clear that there are large differences in the quality of education across countries and regions, which implies that the human capital produced by the same number of years of education in different countries is supposed to be different, and measuring human capital in terms of years of schooling implicitly assumes that the same number of years of schooling across different countries and regions produces the exact same educational outcomes, which is clearly not the case. factually incorrect. For example, few would in fact argue that one year of education in a U.S. secondary school is equivalent to that in Egypt (Hanushek & Kimko, 2000). Based on this view, it is clear that average years of schooling, as a quantitative indicator of a country's education, is not a good reflection of a country's human capital.

While the first criticism diminished as the data improved, the second grew stronger. Researchers have realized that having human capital measures should reflect both the quantity of a country's human capital as well as a good measure of the quality of a country's human capital. In order to control for differences in the quality of education across countries in their studies, a number of studies have used indicators such as teacher-student ratios and per-pupil expenditures as proxies for the quality of education in each country, but these indicators are all elements of the inputs to the production of education, and it would be inappropriate to consider all of these inputs as effective productive investments (Hanushek, 2003), and there is uncertainty about how much of the inputs are able to produce the outputs (Schoellman, 2012; Hanushek, 2003). There is still a great deal of controversy in the academic community as to whether "educational inputs are useful or useless". For example, rural schools have smaller pupil-teacher ratios than urban schools, but the quality of education in rural schools is usually poorer than in urban schools; similarly, small schools have higher per-pupil costs due to scale effects, but the quality of education in small schools is not necessarily higher than in other schools.

Measuring the quality of a country's education and human capital in terms of educational outcomes has great comparative advantages in terms of doctrine. Along with the advocacy that "going to school is not the same as learning" (Pritchett, 2013), "(student) cognitive skills" have stood out from the rest and entered researchers' of researchers. Human capital theory suggests that a person receives education in order to acquire skills, which can lead to an increase in the productivity of one's labor, which in turn contributes to higher personal incomes and national economic growth. Consequently, cognitive skills have been introduced into the measurement of human capital, which is considered a return to the concepts of human capital theory (Hanushek & Woessmann, 2015; Huang, B. et al., 2024). Students' cognitive skills were first used in numerous studies due to the fact that international and regional organizations conduct extensive student competency assessment test items in different countries, which measure a variety of literacies, such as mathematics, literature, and science, and whose scores for each of these literacies provide a good measure of students' cognitive skills. The use of scores from these test programs presupposes that these scores are comparable across test programs, and a number of scholars have worked to address this issue, and has emerged as a number of very valuable methods (Hanushek & Woessmann, 2012; Angrist et al., 2021; Gust et al., 2024).

[6]Despite the growing importance of students' cognitive skills as a measure of education quality, and the fact that studies have found that when controlling for both mean students' cognitive skills and average years of schooling, only the mean students' cognitive skills significantly affect a country's economic growth , average years of schooling as a measure of the quantity of education has not been eliminated altogether. This is partly due to the unrivaled advantage of years of schooling in terms of widespread acceptance, and partly because the academic community has never denied the importance of quantity of education, which is considered to be as important as quality. In this context, a series of studies have emerged that seek to combine the quantity and quality of education by means of methodologies that allow the construction of new indicators that contain information on both the quantity and quality of education. Such studies typically estimate the quality of education from data on students' cognitive skills, calculate a quality-adjusted coefficient with reference to a benchmark (country), adjust the average years of schooling, and obtain quality-adjusted

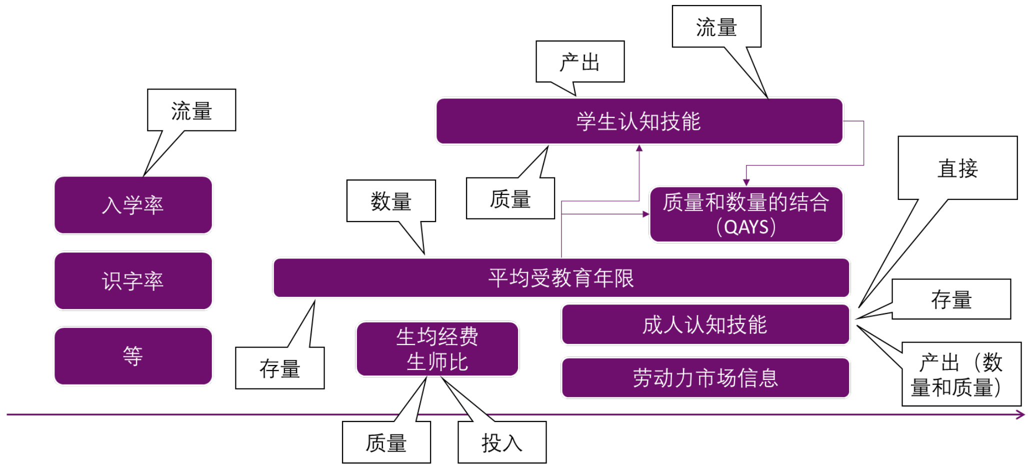
[6] It has been argued that this is due to the large covariance between average years of education and the mean value of students' cognitive skills, rather than the uselessness of the amount of education.

years of schooling (Filmer et al., 2020).

At the same time, too much use of student cognitive skills falls into the same flaw of replacing stock with flow as enrollment[7] . A number of problems arising from the use of student cognitive skills can be solved by the direct use of adult cognitive skills data. In addition, because adult cognitive skills data have the best qualities of both educational output and stock, and because the measurement of cognitive skills can be treated as a direct measure of human capital according to the human capital concept, adult cognitive skills are naturally the optimal choice for human capital measurement. 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 containing more countries by establishing the correlation between students' cognitive skills and adults' cognitive skills based on the relationship between flow and stock, which is a similarly groundbreaking approach (Égert et al., 2024) . Although the method is still in its infancy, it has great application value in the future and is well worth paying attention to.

In addition, there are some scholars who have adopted different approaches in constructing new human capital measures. For example, some studies have taken a different approach by escaping the analytical framework of education input-output altogether and proposing a rather innovative methodology based on the fact that wage information in the labor market is an intuitive reflection of human capital, separating the quality of education and human capital in each country from the wage information (Schoellman, 2012; Martellini et al., 2024). The method is currently deficient in terms of changes in the quality of education and measuring increments, although its value should not be overlooked, due to, among other things, problems of access to wage information in the market.

[7] In addition to the theoretical inappropriateness, there has been empirical evidence that although there is some correlation between the various characteristics of students' and adults' cognitive skills, there is a large difference especially in the characteristics of distributional skewness and standard deviation (Huang, B. et al., 2024). These have prompted scholars to rethink the use of student cognitive skill data as a substitute for adult cognitive skill data in research.



Note: Time advances gradually as the axes move from left to right.

Figure 2-1 Broad lines of development of human capital measurement indicators

Next, this paper will systematically introduce their problems, data construction principles and methods, and existing public data along the development of national human capital, so as to lay the foundation for comparing and analyzing the advantages and disadvantages of various education human capital measurement indicators and summarizing the basic principles for judging the advantages and disadvantages of education human capital.

III. Years of schooling: a measure of educational attainment [8]

As of today, years of schooling remains the most prevalent indicator of human capital, and average years of schooling remains the most commonly used measure of a country's human capital in macro studies. It is widely recognized that average years of schooling is a measure of the quantity of education quality acquired in a country.

(i) Completion of data for missing years: interpolation of years of schooling

In general, the average years of schooling for each country is calculated from individual microdata from national censuses or surveys. There are two problems with the calculation of this indicator:

[8] It should be recalled that in English, Educational Attainment (EA) usually refers to the level of education or years of schooling, and its connotation is inherently skewed towards the quantitative dimension, without much reference to quality.

First, there is a lack of uniformity in the classification of levels and types of education across countries. On the one hand, education systems and schooling systems differ from country to country, with each country having a different education system from the others, encompassing different types of schools, and with some differences in the number of years required to complete what appears to be the same level of education in different countries; on the other hand, even within the same country, schooling systems may change from one period to the next. Addressing this issue requires a careful comparative analysis of the education system and the school system in each of the countries in the sample. This work does not involve complex technology, but it requires a great deal of time and effort to organize due to its tedious complexity.

Second, censuses in countries are usually conducted every 10 years, which is too long an interval, and to produce panel data on years of schooling at shorter intervals (e.g., 5-year intervals) requires data interpolation for the intervening years for which there are no censuses or surveys. Several methods have been proposed by previous scholars to achieve data interpolation. Early studies proposed relatively simple methods, including the Perpetual Inventory Method (Barro & Lee ,1993; Barro & Lee, 2001), Simple Linear Extrapolation (De La Fuente & Doménech, 2000; De La Fuente & Doménech, 2006), etc.[9] . In 2007, researchers began to use population birth cohort information for data interpolation, proposing the Forward and Backward Extrapolation [10] (Cohen & Soto, 2007; Barro & Lee, 2013; De La Fuente & Doménech, 2015; Barro & Lee, 2015) and 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 macro studies today is constructed based on the methodology of Barro & Lee (2013), using birth cohort trend extrapolation. In order to enhance the understanding of mean years of schooling, and taking into account the history of development and applicability of the methods, this paper will present both the perpetual inventory method, the birth cohort trend extrapolation method, and the birth cohort iterative backward extrapolation method, which are (were) used in the mainstream. The principles of these three methods are different, but related: perpetual inventory uses base period data to interpolate to

[9] In contrast, the principle of simple linear interpolation is simpler, and this method will not be described in this paper. For an earlier method of interpolating years of schooling, see De La Fuente & Doménech (2006).

[10] This series of articles doesn't name its methods; we've taken keywords as the names of its methods.

subsequent years, iterative backcasting uses the latest data as the base period to interpolate to previous years, and trend extrapolation is based on both forward and backcasting of base period data.

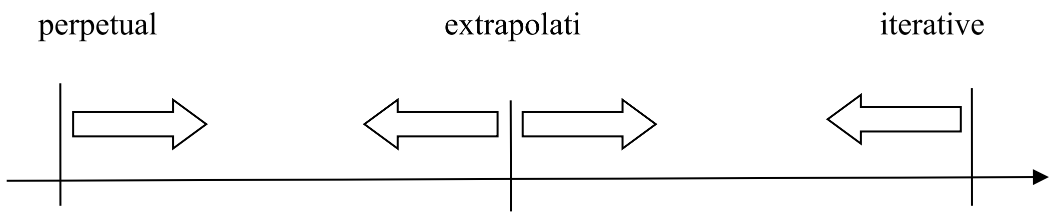


Figure 3-1 Schematic diagram of the three data interpolation methods

(ii) Interpolation of years of schooling

1. Perpetual inventory method

The perpetual inventory method uses census or survey data for a given period in a country as the base period and takes into account changes in the population at all levels of education in subsequent periods due to deaths and new enrolments[11] to this to achieve a complementary value for the years of schooling data for subsequent years in that country.

Typically, the average number of years of schooling for the population aged 15 years and over in each country is obtained by multiplying and summing the proportion of the population aged 15 years and over at each stage of education with the time required to obtain that stage of education, i.e., there:

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

[11] For statistical completeness, there are two types of statistics. The first one is commonly used, and is also used in this paper and the cited literature, which vertically stratifies the population by the highest education level, where the population at each higher education level is the population with the highest education level (e.g., uneducated population, elementary school population, lower secondary school population, etc.), and the corresponding years of education is the time needed to obtain the corresponding education level, e.g., 6 years for elementary school, 9 years for lower secondary school, 12 years for upper secondary school, etc.; the second one is to horizontally stratify the population, where each higher education level is the population that has completed that education level. The second is to stratify the population horizontally, in which the population of each education stage is the population that has completed that education stage (e.g., the uneducated population, the population that has attended elementary school, the population that has attended elementary school and then junior high school, the population that has attended junior high school and then university, etc.), and the corresponding number of years of education is the amount of time needed to obtain the education stage, e.g., six years for primary school, nine years for junior high school and twelve years for senior high school, etc. The second is to stratify the population horizontally. (e.g. 6 years for elementary school, 3 years for lower secondary school, 3 years for upper secondary school, etc.). In fact, when the structure of the educated population is obtained for each country, both statistical methods give consistent results.

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

where ys_t is the average number of years of schooling of the population aged 15 years and over in a given country in the year t ;[12] ; l denotes the stages of education, which are usually classified using a broader hierarchy of levels of education: $l = noedu$ for those with no education, $l = pri$ for those who have completed primary education, $l = sec$ for those who have completed secondary education, and $l = ter$ for those who have completed tertiary education; $h_{l,t}$ denotes the percentage of the population at each stage of education, which is equal to the number of people at each stage of education l ($H_{l,t}$) divided by the total population (L_t); and $Dur_{l,t}$ is the number of years of education required to reach the level of education in the country. is the number of years of education required to reach that level of education in the country.

Assume that we have the number of people in each educational stage in the base period ($H_{l,t-5}$) and want to make up the value of the number of people in each educational stage five years later ($H_{l,t}$). According to the perpetual inventory method, the following formula can be used:

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

where δ_t is the mortality rate, $H_{l,t-5}(1 - \delta_t)$ is the population surviving at each stage of education adjusted for deaths , and $add_{l,t-5\ to\ t}$ is the change in the population added at each stage of education, which is usually measured using information on school enrolment.

This formula can be visualized as the population at each level of education being equal to the number of people who completed that level of education in the base period, minus the number of people who died at that level of education, plus the number of new people who completed that level of education. For example, if a country conducts a census in 2010, the average number of years of schooling of the population in 2010 can be calculated according to the above formula (3-1)-(3-2), but if we need the number of years of schooling of the population of that country in 2015 , we can apply the formula (3-3), i.e., calculate the following formula:

[12] The average years of schooling calculated in the existing literature all begin at age 15 and will not be specifically emphasized later.

$$H_{l,2015} = H_{l,2010}(1 - \delta_{2010}) + add_{l,2010\ to\ 2015}{}^{13} \quad (3\text{-}4)$$

As can be seen from the above equation, mortality rates do not vary with birth cohort (age) and educational stage, which is considered to be the biggest problem with the perpetual inventory method, as the reality tends to be that the higher the level of education, the lower the mortality rate (Balaj et al., 2024); and the higher the age, the higher the mortality rate[14] . It has been suspected that the large errors arising from this shortcoming are likely to be the main reason for the failure of empirical studies to obtain satisfactory results (Krueger & Lindahl, 2001; De La Fuente & Doménech, 2006).

2. Extrapolation of birth cohort trends

Utilizing information from the birth cohort can effectively reduce the problem of measurement error caused by mortality assumptions (Cohen & Soto, 2007). Under the birth cohort perspective, the formula for the average years of schooling of a population is as follows:

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

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

where ys_t is the final total mean years of schooling, ys_t^a is the mean years of schooling for 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 birth cohort at each level of education. The new symbol a is introduced here for the birth cohort, which differs slightly across studies, but usually indicates 15-19 years of age when $a = 1$ is used, 20-24 years of age when $a = 2$ is used, and so on, up to $a = 11$ for 65 years of age and older. Unlike before, here the entire population is expanded into birth cohorts by age, and after obtaining the mean years of schooling for each birth cohort (ys_t^a), the population's mean years of schooling is obtained by weighting and summing the population according to the population share

[13] The most important aspects of the application of the method are the estimation of mortality rates over time δ_t and the use of enrollment information to measure the change in the population of each educational stage added $add_{l,t-5\ to\ t}$, which will not be described in detail in this paper due to space constraints, but can be found in Barro & Lee (2001), if needed, or the authors of this paper may be contacted for specifics of the methods of this literature, summarized by the authors of this paper. details and descriptions.

[14] In addition to this, the onset of migration can also lead to changes in the proportion of the population at each stage of education at different times and for the same birth cohort, which is then only marginally discussed in the existing studies, and no study has yet taken this into account in its methodology.

of each birth cohort (l_t^a) (ys_t) .

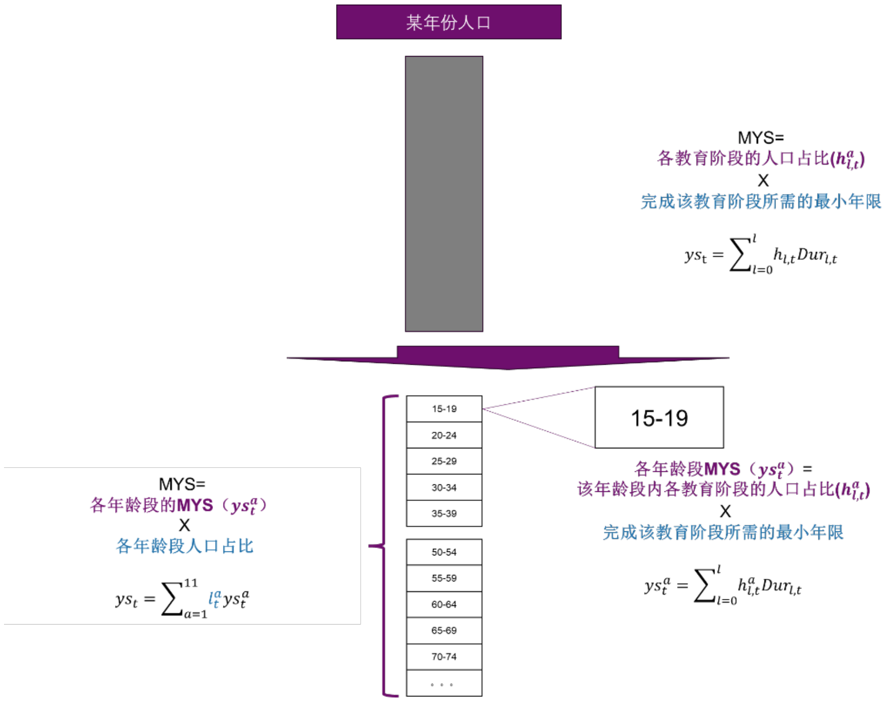


Figure 3-2 Schematic diagram for calculating average years of schooling

In the above equation, the demographic information is available from the World Population Prospects (WPP) of the Population Division of the Economic and Social Affairs Department of the United Nations and is considered as known. Therefore, to obtain the average number of years of schooling in each country, it is only necessary to know either the percentage of the population in each educational stage for each birth cohort in a given country ($h_{l,t}^a$) (Barro & Lee, 2013) or the average number of years of schooling (ys_t^a) (Cohen & Soto, 2007). Both pieces of information can be obtained using birth cohort extrapolation.

Birth cohort extrapolation assumes that for an individual who completes formal education, the level or years of education will not change throughout his or her life, which means that the proportion of the population in the same birth cohort in each educational stage will always remain the same unless the birth cohort of that population undergoes a significant structural change due to mortality (Barro & Lee, 2013). Since an individual's years of schooling remain constant throughout his or her life, and the proportion of the population in the same birth cohort across education stages also remains constant, the average years of schooling for the population in that birth cohort will also remain constant (Cohen & Soto, 2007). Using this inference, we can then forward and backward extrapolate the population's years of schooling data for previous and subsequent years based on the base period data.

It is important to note that the birth cohort extrapolation method also makes an important assumption about population mortality, which assumes that mortality rates do not change over time with educational attainment, i.e., survival rates are the same regardless of educational attainment in the same birth cohort. This assumption is made in order to ensure that the percentage of the population in the same birth cohort at each educational level does not change over time (Barro & Lee, 2013), and subsequently to conclude that the average number of years of schooling for the population in that birth cohort also does not change over time (Cohen & Soto, 2007). Barro & Lee (2013) found from the available Census information, found this assumption to be true for the population aged 64 and under and not for the older cohort (population aged 65+), thus requiring mortality adjustments for the older cohort. In addition, for the younger age groups under 25 years old, since the education status of these individuals is still evolving, other methods of estimation are also needed[15] .

To understand this method more intuitively, we rely on Barro & Lee (2013) and plot Figure 3. In the figure, period t is the base period for which the data are owned, and t+5 and t-5 are the years to be interpolated. In this figure, the method needs to accomplish the following two components:

First, the missing data $h_{l,t\pm5}^{a\pm1}$ or $ys_{t\pm5}^{a\pm1}$ for the period $t\pm5$ can be obtained by having the data t for the periods $h_{l,t}^{a}$ and ys_{t}^{a} by forward and backward pushing. As shown by the solid arrows in the figure, for example, by using $h_{l,t}^{4}$ and ys_{t}^{4} in the birth queue (30-34) of the t period, $h_{l,t-5}^{3}$ and ys_{t-5}^{3} in the birth queue (25-29) of the $t-5$ period, and $h_{l,t+5}^{5}$ and ys_{t+5}^{5} in the birth queue (35-39) of the $t+5$ period, can be obtained by backward and forward extrapolation, respectively. In general, in the forward projection, the following two equations can be utilized to obtain $h_{l,t+5}^{a+1}$ or ys_{t+5}^{a+1}

[15] While sharing the same understanding of this, the formulas of Cohen & Soto (2007) and Barro & Lee (2013) have subtle differences in the birth cohorta. In the forward projection Cohen & Soto (2007) use 25-29 in period t to push 30-34 in period t+5, while Barro & Lee (2013) use 20-24 in period t to push 25-29 in period t+5; in the backward projection Cohen & Soto (2007) use 30-34 in period t to push 25- 29, while Barro & Lee (2013) use 25-29 in period t to push 20-24 in period t-5.

for the missing years:

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

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

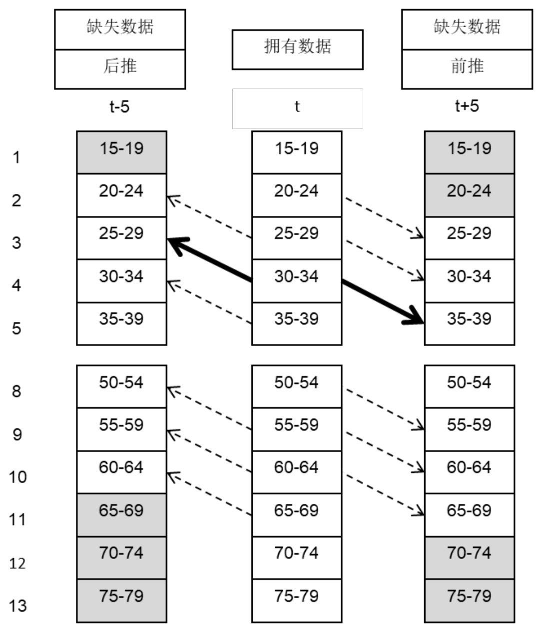
Similarly, in post-projection, the following two equations can be utilized to obtain$h_{l,t-5}^{a-1}$ orys_{t-5}^{a-1} for the missing year:

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

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

All of the above formulas work by equating information from the same birth cohort between years.

Second, birth cohorts for which data are not available through trend extrapolation need to be otherwise topped up. As in the gray underlined birth cohorts in the figure (birth cohorts before age 25 and after age 64), the combined effects of school enrollment, mortality, migration, and other factors are usually taken into account in the value-added treatment.[16] .



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

Figure 3-3 Schematic of the Trend Extrapolation Method

The general steps in this type of study are usually as follows: first, information needs to be obtained by trend extrapolation (i.e., Equation (3-6)-(3-8) or Equation (3-7)-Equation (3-9)) for the birth cohorts for which information is available; second, certain methods are used to estimate the information for the birth cohorts for which information cannot be obtained by trend extrapolation; and finally, Equation (3-4)-(3-5) is used to compute the missing data years for the Average years of schooling.

3. Iterative backcasting of the birth queue[17]

Unlike the birth cohort trend extrapolation method, the birth cohort iterative backpropagation method[18] only performs backpropagation and does not do forward extrapolation. The method usually selects the most recent year of data as the base year, and iteratively backcasts and backfills based on that year's data without stopping. The implementation of the method can be roughly divided into the following two parts:

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

(1) Data on education information for the base year (t), which is usually derived from the census. After harmonized data preprocessing, the population shares by sex-birth cohort-education stage for base year t are obtained$h(a, l, t, sex)$[19] .

(2) Demographic data for all years, consistent with several articles above, this data is from WPP.

(3) Life Table data for the calendar year, also from WPP, which is used to calculate the survival rate for the "gender-birth cohort-education" population for the calendar

---

[17] The data constructed by this method are far less common in applications than those constructed by other methods, and therefore, this paper will only provide a brief introduction to this method. For further information, see Lutz et al. (2007) and Speringer et al. (2019).

[18] The original paper calls this method Iterative Multi-dimensional Cohort-component Reconstruction (Iterative Multi-dimensional Cohort-component Reconstruction)

[19] In this step Lutz et al. (2007) used the data on the number of people by sex - birth cohort - education stage, and the data on the number of people were obtained by using the percentage of people by sex - birth cohort - education stage and the data on the demographic structure of the UN (birth cohort - number of people). Therefore, the formula in step (4) corresponds to the number of population, and after step (5) the number is converted to the percentage of population in combination with the demographic data of UN to obtain the new data of the number of population of each sex-birth cohort-education stage, and this conversion of the number of population-population-percentage-of-population-quantity is artificially considered as a consideration of the migration situation. In Speringer et al. (2019), on the other hand, the sex-by-birth-cohort-education population shares are used from start to finish, and only the final population shares are combined with UN's demographic data to obtain population numbers.

year (e.g.$t-5$) in conjunction with information on the difference in life expectancy by gender-education stage. Survival Ratios$(a-1, l, t-5, sex)$ is used for the survivorship adjustment.

The second step is an iterative computation, and to introduce the iterative process, we use the example of constructing 15-19 to 105+ birth cohorts[20] , a schematic of one complete iteration of the process is given in Figure 2-3, where the text corresponds to the steps and names of the iterations.

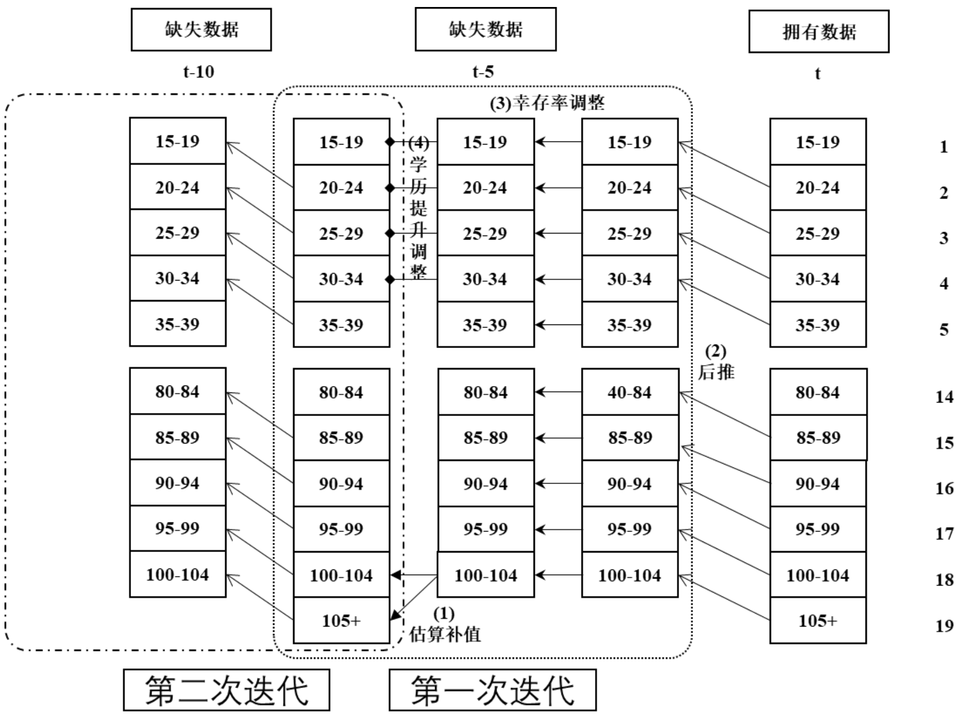


Figure 3-4 Schematic of one complete iteration of

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

(1) Estimate the complementary value for the highest birth queue. As shown in the figure, during the first iteration, since the birth cohort 105+ of the$t$ period is used to back-propagate the birth cohort 100-104 of the$t-5$ period, at this point for the$t-5$ period, its highest birth cohort (105+) is vacant, and if a second iteration is to be performed, this information needs to be topped up first. Since it is only after the iteration that the highest age cohort data becomes vacant and needs to be topped up, this step needs to be considered only during the second and subsequent iterations.

[20] 70+ in Lutz et al. (2007); 100+ in Goujon et al. (2016); and 105+ in Speringer et al. (2019).

(2) Backward extrapolation. This step is similar to the trend extrapolation method, except that the following formula is used to obtain data on the population share of "different sexes - different birth cohorts - different stages of education" for each birth cohort in period t-5:

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

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

(3) Survival adjustment. Deaths are changing the demographics of the population from$t$ to$t-5$, so a survivorship adjustment needs to be made to the extrapolated results using survivorship informationSurvival ratios$(a-1, l, t-5, sex)$:

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

(4) Adjustment for educational upgrading. Since the possibility of educational upgrading still exists for the 15-34 year old population, for this reason an adjustment for years of schooling is applied to the 15-34 year old cohort obtained by backcasting. Obtain the final$t-5$ period results.

(5) Going back to step 1, a second iteration of the calculation is performed based on the results of period t-5, and so on to obtain the results of period$t-10,\ t-15, \dots,$.

The above (2) and (3) can be synthesized into one step. In the application of the method, the literature basically follows the above steps for iterative complementary values, although the stage treatment of each part is slightly different, especially in steps (2), (3) and (5).

(iii) Literature summary and introduction to the average years of schooling database[21]

1. Summary of literature

Table 3-1 Average Years of Education Methodology-Literature-Database

[21] We have presented three different basic methods above, and then the application of these methods across the literature is far more complex and requires consideration of various details. Due to space constraints and because the specific application of the methods remains a technical part, with too many, indispensable formulas that increase the difficulty of understanding, this paper does not present the specific details of these methods, which are different in each article. Readers who are interested can read the original article or can contact the authors of this paper for specific details and descriptions of the studies used for the above basic methods,

Correspondence Table

methodologies	literatures
perpetual inventory method	Barro & Lee (1993) Barro & Lee (2001)
Extrapolation of birth cohort trends	Cohen & Soto (2007) Barro & Lee (2013) Barro & Lee (2015) De La Fuente & Doménech (2015)
Iterative backpropagation of birth queues	Lutz et al. (2007) Bauer et al. (2012) Goujon et al. (2016) Speringer et al. (2019)

2. Introduction to the database

The perpetual inventory methodology is an early use of the methodology, and the data constructed are no longer favoured by existing research, and there is no longer an organization or individual to update the original data. Currently publicly available, continuously updated databases include the Barro-Lee Educational Attainment Dataset and the Wittgenstein Center Human Capital Data, which are based on trend extrapolation and iterative backward extrapolation, respectively. Capital Data.)

Among these, the Barro-Lee Educational Achievement Database is the most commonly used source of data on average years of schooling in current research, constructed according to Barro & Lee (2013). The most recent data available is the 2021 update, which provides panel data on average years of schooling by age and gender for 146 countries from 1950 to 2015.

In addition, the Wittgenstein Center Human Capital Data also provides panel data on the distribution of educational attainment by age and gender for 185 countries for the period 1950-2015, along with panel data on the average number of years of schooling based on educational attainment. Comparatively, the Wittgenstein Center Human Capital Data provides more countries, more content, and longer age ranges, but its use in the literature is currently less frequent and far less accepted than the Barro-Lee Educational Attainment Database. We speculate that the excessive smoothing, estimation, and prediction in the data construction process of the Wittgenstein Center human capital data is the main reason for its limited data use.

---

summarized by the authors of this paper (Appendix I). The same article follows.

Table 2-2 Average years of schooling database

	Barro-Lee Educational Achievement Database	Wittgenstein Center Human Capital Data
Usage	Trend extrapolation (forward+ backward)	iterative backcasting
Methodological literature	Barro & Lee (2013)	Speringer et al. (2019)
nations	146	185
timing	1950-2015	1950-2015 (2020 data forthcoming)
Last Updated	2021	2024
Contents	Average years of schooling	Distribution of access to education, average years of schooling
age groups	15-24, 25-34, ... 55-64 years old 15-64; 25-64	15-24, 25-34, ... 100+ years old 15+; 25+, etc.
distinguishing between the sexes	Male and female; male and female	Male and female; male and female
website	http://barrolee.com/ https://github.com/barrolee/BarroLeeDataSet	https://dataexplorer.wittgensteincentre.org/
acceptability	your (honorific)	lower (one's head)

IV. Students' cognitive skills: a measure focusing on the quality of educational acquisition [22]

Student cognitive skills are an indicator that has only emerged in the last two decades, and mean student cognitive skills are increasingly being used in macro-level studies, and other statistics on the distribution of student cognitive skills, while not a priority, are also well covered in the literature. Student cognitive skills are often used as a measure of the quality of educational attainment in macro studies, and in fact there is little discussion of whether they are a measure of educational quality in their own right, as we will discuss in Chapter Eight.

(i) Construction of an internationally comparable database: comparability of scores across test items

Over the past two decades, the importance of the quality of education has become increasingly evident as the academic community has recognized and emphasized that

[22] The term quality has different meanings in different scenarios. For example, if an education system has five years of education and each year of its education improves its math score by 20 points, then ultimately the quality of this student is 100 points and the quality of each year's education is 20 points, with the former being the overall, final quality and the latter being the unit quality. Although student cognitive skills are a total output, since they are unit quality at a certain student age or grade level.

"schooling is not learning". As a result, more and more international or regional organizations are conducting assessments of student literacy in basic education to evaluate the quality of education in various countries, such as the Program for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), International Reading Study (IRS), International Reading Council (IRC), and the International Reading Council (IRC). Program 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). Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), Programme d'Analyse des Systèmes Éducatifs de la francophonie (Program for the Analysis of Education Systems in French-speaking Countries. (PASEC), the First, Second and Third Regional Comparative Studies (First/Second and Third Regional Comparative Studies) carried out by the Latin American Laboratory for Assessment of the Quality of Education (LLECE), and the National Education System Analysis Project (NESAP), which was launched at the end of 2009 in the French-speaking countries. (PERCE/SERCE/TERCE), the Early Grade Reading Assessment (EGRA), the Annual Status of Education Report (ASER), and the National Education Report (NER). (EGRA), Annual Status of Education Report (ASER), United States' National Assessment of Educational Progress (NAEP), India National Achievement Survey (NAS), and the National Achievement Survey (NAS). India National Achievement Survey (NAS)[23] . These programs test a variety of cognitive skills that can influence students' future productive behaviors, such as mathematical literacy, literacy (reading and writing), scientific literacy, etc. Scores on these skills and competencies provide a good measure of student learning gains in each country, which can be used to further evaluate the quality of each country's education system at that time.

Student cognitive skill scores have been recognized as the best measure of educational quality in a number of macro-level studies. However, studies often want to have data from more countries (cross-sectional) and over a longer period of time (longitudinal), and the various test items are not comparable by design, which requires methods to make the test items cross-sectionally and longitudinally comparable.

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

Specifically, in the cross-section, the study wanted to make the two test programs comparable for similar years[24] . The aim is twofold: to include a larger number of countries, especially at all stages of development and in all regions. However, the number of countries included in the individual tests is quite limited - even PISA, which covers the largest number of countries, surveys only 102 countries and territories, which is less than half the number of countries in the world; in terms of the level of economic development, these countries are more likely to be upper-middle-income and higher economies, with less coverage of lower-middle-income and lower-income economies; and regionally, the number of countries in the African region is relatively small. Regionally, the African region has fewer countries. The solution to this problem lies mainly in making the scores of the various tests comparable, so as to include a larger sample of countries by integrating the international and regional tests, which is the main objective of the existing research. The second is to integrate the different information between the two tests to enable cross-regional comparisons of a given piece of information. Since this purpose is currently found in only one type of study, we will present details later (see 4.2.2).

Longitudinally, macro studies also often want to include more years of data (cross-sectionally), which means that some methodology is needed to make scores comparable across multiple rounds of testing. After the 1990s, tests were largely based on item response theory, whereby the same questions (see 4.2.3) made the tests themselves longitudinally comparable over time, and therefore did not require additional processing. For the pre-1990s tests, on the other hand, which amounted to treating the different rounds of the same test program as the results of two completely different test programs, the treatment was the same as in the cross-sectional countries, except that it took into account possible differences in the overall level (mean) due to time (from this point of view, the cross-sectional and longitudinal comparability was the same). Given the methodological similarities and the specificity and sparseness of use of the data prior to the 1990s, the temporal comparability will not be presented too much in this paper.

[24] The conversion of multiple different test items is also done in two and two, which are used throughout the text for precision of presentation.

(ii) Anchor points and constructor conversion functions

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

Suppose you want to convert the score of test item X to the score of test item Y. The generalized formula is:

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

Where $Score_Y$ is the score of the test item Y that we want to get, $Score_X$ is the score of the test item X that we already have, and $f(\cdot)$ is a function that, if simple linearity is used, the formula can be changed to:

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

Where α and β are conversion parameters.

Using conversion functions, we can achieve both micro-level conversions of individual student scores and macro-level conversions of national mean or total scores, depending on the methodology and level of data used.

The key to constructing a transformation function is to have anchoring points, which are places of overlap between two different test items. Anchoring points can be students, countries, and topics (or questions) that overlap between two tests. Only by having anchor points can conversion formulas be constructed; it can be argued that the same individuals (Common Persons), the same populations (Common Populations), or the same items (Common Items or Overlapping Items) are the basis for constructing conversion functions (Kolen & Brennan, 2014; Reardon et al., 2021)[25] . When we apply different types of anchors, the techniques used will vary, and the following figure gives the overall framework of the approach, which we will describe one by one.

[25] In this article, common, overlapping, same, and repetitive represent the same meaning.

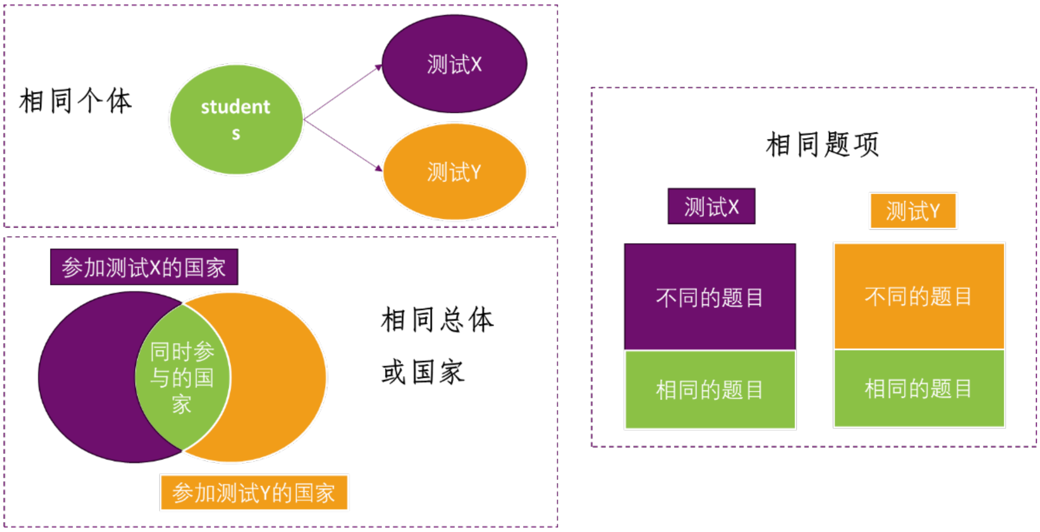
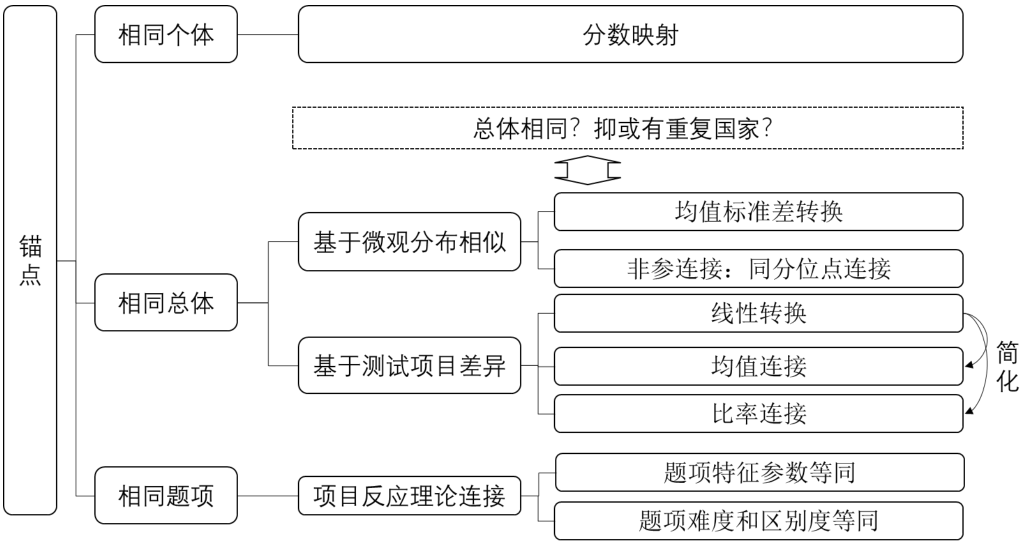


Figure 4-1 Schematic diagram of the anchor point situation

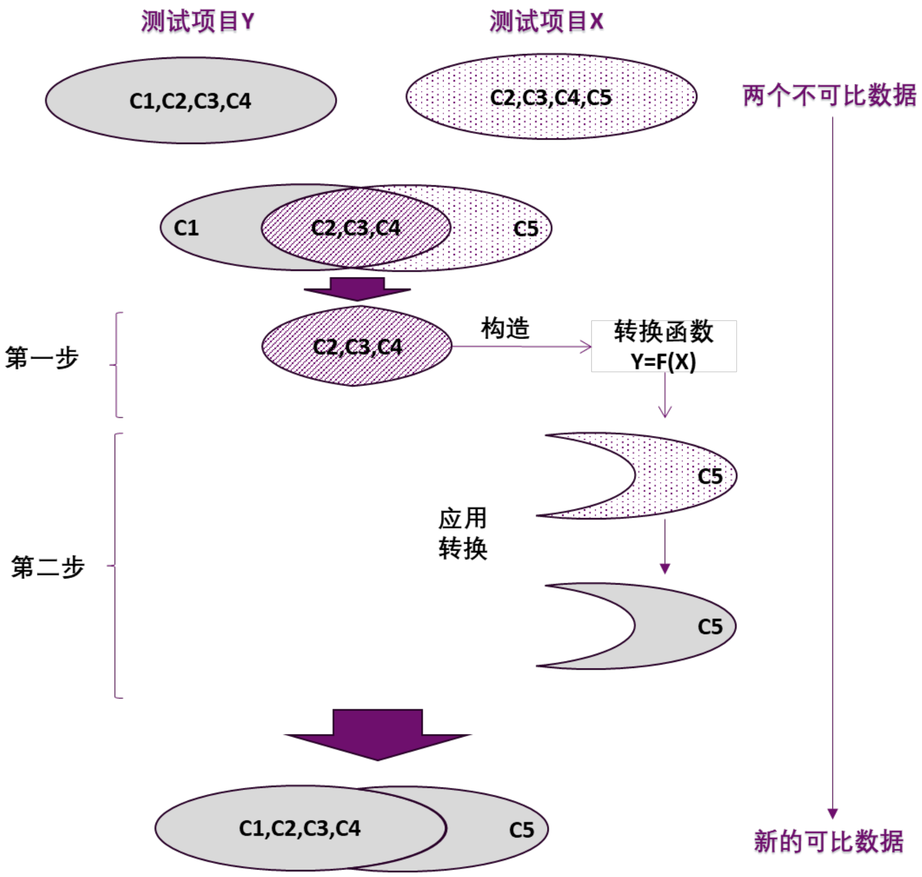


Note: Self-painted by the author.

Figure 4-2 Methodological framework for constructing transformation relationships between different test identities

In addition, regardless of the purpose, as mentioned earlier, this type of research is usually divided into two steps: in the first step, a function (instrument) is constructed using anchors, and in the second step, the function (instrument) is applied to realize the transformation. In the figure below, using the same country (see 4-2-2) for the purpose of expanding the sample of countries as an example, it is assumed that there are two test items X and Y (indicated by different underlining shading in the figure), both of which conduct tests of students' mathematical literacy but cover different countries and regions, where test item X covers countries C1, C2, C3, and C4, and test item Y covers

C2, C3, C4, and C4. where C2 , C3, and C4 participate in both tests (as shown in the shaded portion of the cross in the figure), while C1 and C2 both participate in one test. To expand the sample of countries, we first construct a conversion function that expresses the quantitative relationship between the scores of the two tests through the sample of countries that participate in both tests X and Y (C2, C3, and C4), and then convert the scores of the countries that participate in only test item Y to the scores of test item X, or the scores of the countries that participate in only test item X to the scores of test item Y, through this conversion function. Using the X to Y conversion as an example, we will get comparable data for five countries (C1, C2, C3, C4, and C5).



Note: Take X to Y as an example.

Figure 4-3 Schematic diagram of the steps of the study (using the same countries as an example)

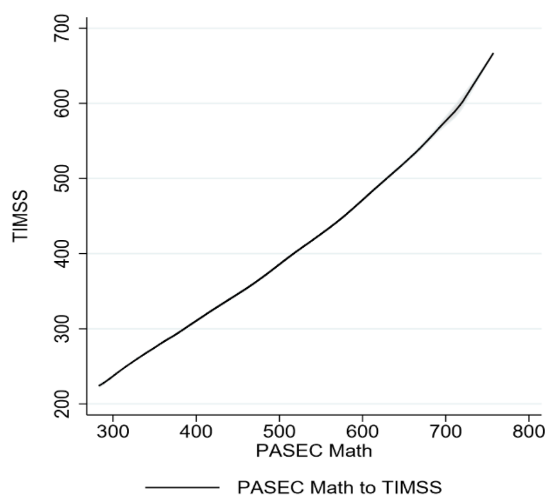
1. Using the same individuals as anchors

When two test items are taken by the same student population and test the same student ability, the difference between the test items will be directly reflected in the different scores of the same students on the two test items because of the stability of student ability. At this point, the same student group serves as an anchor point to directly construct a mapping function between the scores of two different test items by using local linear regression through the students' scores on the two test items. This is represented by the formula:

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

The mapping relationship indicates that each score on test item$X$ corresponds to a score on test item$Y$. After obtaining the mapping relationship, it can then be used for purposes such as expanding the sample of countries. For example, if there is a country where students only took test item X and did not take test item Y, we can get the score of each student in that country if they took test item Y based on the table of score correspondences for X Y.→

For example, Patel & Sandefur (2020) assembled a group of students in India and subjected them to both the PASEC and TIMSS tests at the same time , and then used local linear regression to estimate the data on the PASEC and TIMSS test scores of these students and constructed a mapping relationship between the scores on these two tests. As shown in the figure, it is a curve that reflects the mapping relationship between PASEC and TIMSS test scores, where there is a one-to-one correspondence between PASEC and TIMSS test scores.



Note: Derived from Patel & Sandefur (2020).

Figure 4-4 Mapping relationship expressed as a two-dimensional curve

Once the mapping relationships are obtained, they can be used for purposes such as expanding the sample of countries. For example, if there is a country c where students only take the PASEC test and not the TIMSS test, we can use the score correspondence table constructed by Patel & Sandefur (2020) to convert the scores of students in country c to scores on the TIMSS test and vice versa.

The benefit of using the individual as an anchor is that it is a transformation of individual student microdata, so the data obtained after the transformation remains microdata at the individual student level rather than summed data at the national or regional level, and the researcher can use the student microdata to do a more detailed analysis and realize the research intent that cannot be achieved at the national or regional macro level. Moreover, after obtaining the mapping relationship, this mapping relationship can be used for later tests due to the stability of the international and regional test items, i.e., they are comparable from year to year.

However, there are some drawbacks to this approach. First, it is doubtful whether the mapping relationships of scores on different tests obtained using students from one or a few countries as anchors can be applied to students from other countries, as the group of students may be more adept at one test and less adept at another, biasing the table of score correspondences obtained. Second, constructing mapping relationships using individuals as anchors usually requires the researcher to personally administer a test, which is more labor-intensive, material-intensive, and financial-intensive, and more costly overall, and only Patel & Sandefur (2020) have used this approach in the literature so far. Thirdly, if the validity of the mapping relationship is to be guaranteed, the entire process of test administration, such as the composition of the test questions, student selection (to ensure that there are students on each score band), needs to be strictly guaranteed to be standardized and scientifically credible, which is undoubtedly harder to achieve.

2. Anchored by the same aggregate (country)

We first need to distinguish between two concepts, the same population and the same countries. When the countries participating in both tests are the same (although the students are not the same due to sampling problems), then the population is homogeneous. In reality, however, it is often the case that only a subset of countries participate in both tests, and studies often use these subsets of countries that participate in both tests as anchors (we call them homogeneous countries to distinguish them from each other).

To make it easier to understand the difference between the two cases, let us assume that there are two test items X and Y. When test items X and Y both measure the same countries C1, C2, and C3, then the aggregate formed by the three countries is the same aggregate using the three countries as anchors, and when test item X measures countries C1 and C2, and test item Y measures countries C2 and C3, then C2 serves as an anchor and is the same country.

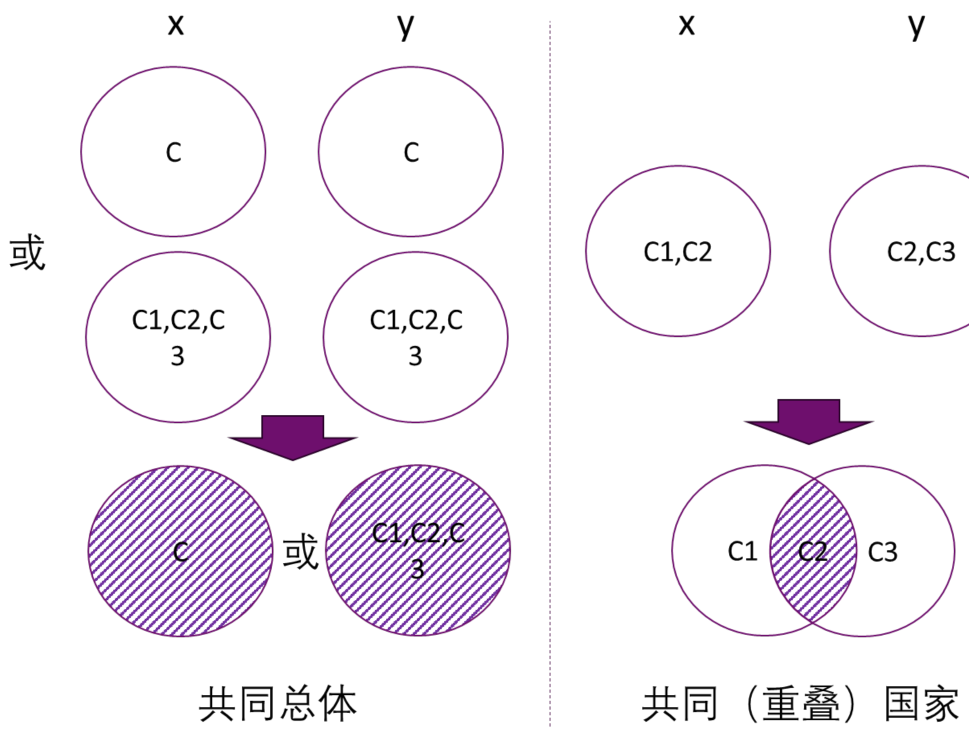


Figure 4-5 Schematic diagram of the distinction between the common aggregate and the common country

The two cases actually reflect two different purposes. For the same aggregate, the purpose is to integrate information from each of the two tests. For example, Test X is implemented uniformly, and the scores of three countries, C1, C2 and C3, are comparable, but there is a lack of information about the provinces (states) within the country; Test Y is implemented by each country individually, and the scores of three countries, C1, C2 and C3, are incomparable, but each country differentiates between the provinces (states) within the country when it implements the test, and it possesses the comparable information about the provinces (states), so that the comparability of provinces (states) can be realized through the three countries of Test X and Test Y, each acting as the same overall. countries each as the same overall, integrating the information from each of the two tests, cross-country comparisons of provinces (states) can be realized.

For the same countries, the aim is to extend the number of countries in the sample. For example, by using country C2 as the anchor country, we can use to determine the relationship of the transformation function between the two test items X and Y. Using this relationship, we can predict the score of country C1, which only participates in test item X, if it participates in test item Y, as well as the score of country C3, which only participates in test item Y, if it participates in test item X. By doing so, we can expand the sample size from having comparable data for only two countries to We can then expand from having comparable data for only two countries to having comparable data for three countries.

The same overall (national) approach needs to be based on two assumptions: the assumption of representativeness, i.e., that individual students in each country participating in both tests must test the same underlying population; and the assumption of content similarity, i.e., that the skills measured by both tests are the same or similar (Tests should measure similar proficiencies). The first is the assumption of representativeness, that is, each individual student in each country participating in the two tests must test the same underlying population, and the second is the assumption of content similarity, that is, the skills measured by the two tests should measure similar proficiencies, as in the case of the PISA and TIMMS tests, which measure mathematical literacy despite differences in design[26] .

In practical research, the largest number of methods are based on the same aggregate (country). These methods are categorized into two main types according to the core assumptions:

The central assumption of the first is the micro-distribution similarity assumption. This is the hypothesis about the distribution of student scores in the same aggregate

[26] What is being expressed here is that, although both PISA and TIMSS are math data, the differences in the topics make them measure differently, and there are many sub-topics under math, and their sub-topics may not be the same, for example, one may be more oriented toward algebraic computational skills and the other more toward spatial geometry skills.

(country) across two different test items, which assumes that if the different test items are accurate measures of the distribution of cognitive skills of the aggregate students in each country, the shape of the distribution of scores of all the students in the same aggregate (country) across the different test items will be the same[27] , differing only in the mean and standard deviation of the distributions. Since this kind of assumption is based on distributions composed of micro-individuals, it is possible to obtain data on individual student micro-individuals[28] . Methods based on this assumption mainly include mean standard deviation transformation and Equipercentile Linking, the latter of which is the only non-parametric linking method in the same aggregate (country) approach.

The central assumption of the second is the test item difference assumption. This is the assumption about differences in scores (means) of the same countries, which argues that systematic differences in the scores of countries in the same countries on the two test items should come from the differences in the two test items rather than from the differences in the countries. Unlike the first hypothesis, the approach based on the second hypothesis emphasizes differences in the aggregate and therefore uses country means to construct the conversion function, so this approach obtains aggregate data at the country level and does not have access to data at the individual micro level. Moreover, such studies are designed to use the two tested replicated countries as anchors, and the purpose of the studies is to expand the sample countries. Methods based on this assumption include mainly Line Linking and, to qualify it, Mean Linking and Ratio Linking.

But either assumption, based on the same countries is also estimating the overall picture using overlapping countries.

As shown in the table below, almost all of the methods in this section have been used to expand the number of countries in the sample in this paper, but only the mean

[27] Since the test items are all measured according to item response theory, more specifically, it should be the case that both test item score distributions are positively distributed. However, all methods really only require the two test distributions to be the same, not necessarily both positive.

[28] It is important to note that in Hanushek's earlier series of studies (Hanushek & Kimko, 2000, Hanushek & Woessmann, 2012a, Hanushek & Woessmann, 2012b), while it is also based on the micro-distribution similarity assumption of the mean- standard deviation transformation idea, but it uses the country's overall mean rather than individual student microdata. And in non-Hanushek studies that use this approach, it is also the country's overall mean that is used, as in Altinok et al. (2018). In effect, these studies temporally treat each country as an individual and use the distribution formed by these individual countries. It is only in Hanushek's recent studies that he applies this method to micro-individual data, e.g. Hanushek & Woessmann (2015), Gust et al. (2024).

standard deviation method is currently being used to integrate information from each of the two tests.

Table 4-1 Objective-Methodology-Assumption Correspondence Table

Same overall: consolidation of information	Mean standard deviation conversion	Similar microdistribution
Same countries: number of extensions	isoform linkage (computing)	
	linear transformation	Differences in test items
	mean value connection	
	ratiometric connectivity	

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

Table 4-2 Correspondence of method names in this paper with other literature

Method naming in this paper		Naming in other literature
Similar microdistribution	Mean standard deviation conversion	Linear connections (Altinok et al., 2018; Angrist et al., 2021)
Differences in test items	isoform linkage (computing)	Homologous site linkage (Altinok et al., 2018)
	linear transformation	Regression (Angrist et al., 2021)
	mean value connection	Mean value connection (Altinok et al., 2018)
	ratiometric connectivity	Pseudo-linear connections (Altinok et al., 2018); ratiometric connections (Patrinos & Angrist, 2018)

Note: Author's own production.

(1) Microdistribution similarity hypothesis: mean standard deviation transformation

Under the first assumption of similarity of micro-distributions, since the shapes of the distributions of the two test items are the same and differ only in the mean and standard deviation of the distributions, the following transformation function can be constructed by simply adjusting the mean and standard deviation of the two distributions using the formula for transformation between distributions, [29] :

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

where $Score_{Yi}$ and $Score_{Xi}$ are the scores of individual i on test items X and Y respectively[30] ; μ_X and σ_X are the mean and standard deviation of the scores of students from anchor countries on test item X, and μ_Y and σ_Y are the mean and standard deviation of the scores of students from anchor countries on test item Y.

Currently, the only one that can be fully counted as identical overall is Reardon et al. (2021), whose intent is to compare school district scores across states, but the nationally standardized NAEP testing program, while comparable across states, does not have information on school districts, while the state-organized testing program, while not comparable across states, has information on scores of districts within the continent. At this point, the same aggregate is the individual states (States) that also participate in the national testing program and also participate in their own testing programs, and by integrating the information from the two tests, cross-state comparisons of school districts can be achieved.

In most studies, it is usually the same countries: in reality, it is often the case that only a fraction of the countries participate in both testing programs (in this case, we call them the same countries), e.g., in the case of international student testing programs, it is usually the case that some countries participate in both TIMSS; some countries participate only in TIMSS but not in PISA; and some countries participate only in PISA and not in TIMSS.In the absence of a strictly identical population, the study is at this point only able to use as many identical countries as possible in the two testing programs C as an anchor point, use the means and standard deviations of the distribution of scores for all students in these identical countries to infer the means and standard deviations of the overall student distribution (Gust et al., 2024), and transform the transformation function:

[29] This formula is analogous to normalizing a distribution and then inverse normalizing it, whereas normalization only adjusts the scale of the distribution and does not change the original information of the data.

[30] Morphing the formula yields: $Score_{Yi} = \frac{\sigma_Y}{\sigma_X} * Score_{Xi} + \mu_Y - \frac{\mu_X}{\sigma_X}\mu_Y$, so the method is also made linear transformation by Altinok et al. (2018). Since this method actually uses the assumption of similarity of micro-distributions, we call it a mean-scaled brick difference transformation rather than a linear transformation. Even when this method is applied to macro-country level data, as in Hanusheck's series of studies, it is still based on distributions, just changed from micro-individual distributions to macro-individual analysis.

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

where μ_X^C , σ_X^C , μ_Y^C , and σ_Y^C are the means and standard deviations of the samples consisting of individual students from the replicated countries on the two test items, and $Score_{PISA,i}$ and $Score_{TIMSS,i}$ are the scores of the students i in PISA and the corresponding scores that can be obtained in TIMSS, respectively. Procedurally, such studies are usually divided into two steps, with the first step using the duplicate countries to construct the conversion function, and the second step using the conversion function to convert student scores from countries that participate only in a particular test item to scores on the target test.

As can be seen from the formula, the above transformation in fact consists of a two-fold transformation: a level transformation (or level adjustment) from the mean and a difference transformation (or difference adjustment) from the standard deviation, which correspond to Difficulty and Discrimination in Item Response Theory (IRT)[31] respectively. Level conversion addresses the problem of identical individuals scoring differently on different test items due to differences in the difficulty of the two test items, and it is assumed here that the difference due to difficulty is the same for all individuals (when the standard deviation is the same, this difference is $\mu_X - \mu_Y$); Differential conversion addresses the problem of two identical individuals scoring differently on different test items due to differences in the discrimination of the two test items. problem, for example, two individuals who have a difference of 10 points in one test item have a difference of only 2 points in another test item.

Since the distributions of the two test programs have the same shape, students' position in the overall population will remain the same through the transformation. However, since there will be some differences in the content of the two test items, this makes the same content assumption not strictly satisfied, such as TIMSS focuses on school curriculum content and PISA focuses on real-world problems (Hanushek & Woessmann, 2012a), which is reflected in the differences in the test questions, making the mathematical literacy of the two tests in terms of connotation There are some differences. In this case, the shape of the distribution of student scores on the two tests will be somewhat different. While it is often unavoidable that the position of individual

[31] See the same questions section for item response theory content.

students in the distribution will change slightly when using distribution transformations, the large correlation between the tests (e.g., between scores on the TIMSS math test and the PISA math test) makes it possible for errors arising from the assumption of identical content not to be strictly satisfied to be within acceptable limits. Thus, while this approach can yield data on individual student microcosms, studies often reduce the impact of this error by shifting the focus to the mean of the distribution rather than to a particular student's score; caution should be exercised if additional information on the transformed microcosms of individual students is used.

In addition, under strict data constraints, it is possible to take two tests with only one country in the same program, and then only one country can be used as an "aiming point" for the transformation (Angrist et al., 2021). The essence of the mean-standard deviation transformation is to use the distribution of individual students in the anchor country to infer the distribution of the potential population, and it is clear that using a sample of students from a single country to infer the distribution of student scores in the population is less plausible, and therefore using a single country as the anchor for the transformation is likely to be subject to large errors.[32] ; in fact, if the number of countries used for the linkage is small, the confidence in the results of the transformation becomes lower when the number of countries used for the linkage is small. In fact, if the number of countries used for linking is small, the credibility of the results using this method becomes questionable (Gust et al. 2024).

(2) Microdistribution Similarity Hypothesis: Equipercentile Linking

At the student level, Equipercentile Linking is a common, nonparticipant method of comparing test item scores that does not utilize item response theory (Kolen & Brennan, 2014), a method developed by Braun & Holland (1982).

Under the assumption of similarity of microdistributions, the shape of the distribution of scores on different test items will be the same for all students in the same aggregate (country), and then it is straightforward to connect the two distributions using congruent quantile equivalence.

[32] Due to the differences in the content of the two tests, it is possible that some students in a given country are better at test X and others are better at test Y, thus making the shape of the two test distributions in the same country inconsistent, in which case the two distributions will be closer to a similarly shaped nontrivial distribution only by utilizing a large number of individuals from more countries.

The methodology usually consists of two elements:

One is the value of the interquartile points. That is, based on the quantile points of the score distribution, the corresponding scores are connected, more intuitively as shown in the figure below:

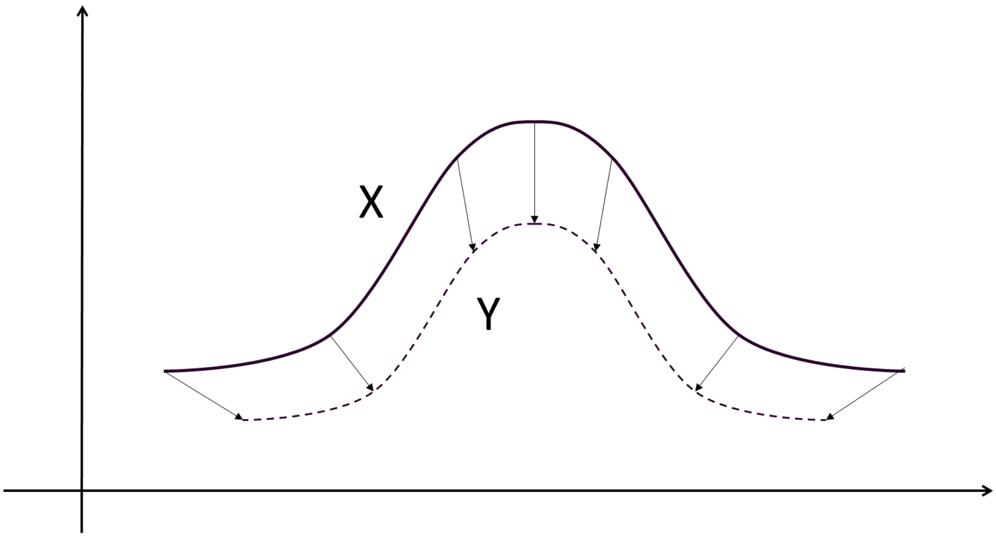


Figure 4-6 Schematic diagram of isochore linkage

There are two achievement distributions X and Y in the graph, and the values (i.e., scores) at no quartile (12th percentile) of X correspond to the values at the same quartile (12th percentile) of Y. The graph can be expressed as the following equation. If expressed in a formula, this can be written as the following equation:

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

The second is smoothing. In the same quartile connection, because the scores are more discrete, we may not be able to find a certain score or a quartile score corresponding to another distribution of scores and quartile scores, so we need to smooth the processing. For example, the 100th and 105th percentile scores of X correspond to the 103rd and 110th percentile scores of Y, respectively, but because the scores are taken more discretely, we do not know how many 102nd percentile scores of X correspond to how many scores of Y. Another example is that, ideally, the 47th percentile score of X should correspond to the 47th percentile score of Y. However, because the scores are taken more discretely, it is made that the closest scores in practice may respectively correspond to X's 47.2th percentile and Y's 47.6th percentile, which is not precise enough, although we can get a rough match.

In this case, while we can utilize Percentile Ranks, the precision 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 error, obtaining empirical distributions and quantile linkage relationships that can most closely approximate the potential aggregate (Altinok et al., 2018).

According to the two smoothing treatments, Equipercentile Linking can be categorized into the following two types: Pre-smoothed Equipercentile Linking (Presmoothed Equipercentile Linking) and Postsmoothing Equipercentile Linking (Postsmoothing Equipercentile Linking). In Presmoothed Equipercentile Linking (smoothing followed by equipercentile linking), the score distribution is usually smoothed using Polynomial Loglinear first, however, the values are taken at the quantile points to construct the correspondence table (Holland and Thayer, 2000); in Postsmoothed Equipercentile Linking (equipercentile linking followed by smoothing), the score distribution is utilized first to perform equipercentile equivalence, and then to construct the correspondence table. equilibrium, followed by smoothing using cubic-spline (Kolen, 1984).

In fact, since it is based on loci, the method does not strictly require that the shape of the distributions of the two test items be the same, but it does require that the same individuals are in the same distributional position in the two test items. This is equivalent to relaxing the assumptions. Therefore, this method is most suitable to be applied when the difficulty of the two test items varies nonlinearly (Altinok et al., 2018).

As before, the goal is usually to convert individual student scores from countries that participated in only one test program to scores from the other test program based on countries that participated in both test programs. The basic approach of this method is therefore to use the distribution of student scores in the same countries to infer the overall distribution of scores, then to construct a correspondence between the distributions of the two test items, and finally to convert the student-level data from the countries that participated in only one test item. And, again, the number of identical countries is an important factor affecting the quality of the linkage. This is because only if the number of identical countries is large enough, the sample of students is large enough for the distribution of the two tests to be close to the overall distribution; and the different quartiles have large enough samples to be estimated. Literature using this approach is mainly Altinok et al., (2018), Sandefur (2018).

(3) Assumption of test item differences: Linear Transforming (Liner Transforming)

The systematic differences in the performance of countries in the same country analyzed in the two test items should come from the differences in the two test items, not from the country differences. Therefore, the transformation parameters can be obtained directly by using OLS to estimate the following equation:

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

Where$\mu_{X,c}$ and$\mu_{Y,c}$ are the mean of the scores of the anchor country$c$ on test$X$ and test$Y$, respectively,α andβ are used by to capture the systematic differences between the two test items [33] . Like the mean-standard deviation transformations, α and β approximately correspond to level-adjusted and difference-adjusted distributional transformations, respectively, reflecting differences in difficulty and discrimination between the two test items. The$c$ in the model denotes those countries that participated in two different test items in the same round of the test program, again using common countries to estimate the aggregate. After constructing the transformation function using the same countries, it is possible to transform the aggregate scores of countries participating only in test X into the aggregate scores under test Y. The transformation function is then used to estimate the aggregate scores under test Y.

Note that after the 1990s, comparability between rounds was achieved by identical question items (see description below) across tests, which also means that the difference between two test items will remain constant across rounds. Therefore, the above equation can be estimated using data from multiple rounds (r indicates the number of rounds):

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

The method of estimating conversion parameters by regression has only recently been used (Angrist et al., 2021), and while the method is again limited by the sample size of the common country, more and more samples will become available for estimation over time, as the testing program continues, and the accuracy of the

[33] The method is not consistent with the linear transformation in Altinok et al. (2018) because the coefficient of β estimation in that method is$\frac{cov(X,Y)}{var(X)}$, while in Altinok et al. (2018) it is$\frac{\sigma(Y)}{\sigma(X)}$, and the relationship between the two is;$\beta = \rho * \frac{\sigma(Y)}{\sigma(X)}$, whereρ is the correlation coefficient. However, it is undeniable that the mean-standard deviation transformation can also be deduced from the mean-connection and ratio-connection, which is exactly the order of the lines in Altinok et al. (2018).

estimated conversion coefficients will increase. However, the problem with this method is also that with each additional round of testing, the parameters estimated by the method will be slightly different.

(4) Testing the hypothesis of item variance: Mean Linking

If the same country participates in two test programs (X ,Y) and receives two scores ($Score_X$ and$Score_Y$), the relationship between these two scores can be given by the following equation:

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

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

$$Score_{Y,c} = \alpha + Score_{X,c} + \varepsilon_c \quad (4\text{-}10)$$

The constantα can be estimated using the same countries participating in both test programs[34] :

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

where$\mu(Y)$ and$\mu(X)$ are the mean averages of the same countries across the two test items. Due to the limitations of the field of view , the authors have not yet found literature on the use of this method for post-1990s test items, but this method, like linear transformation, allows for the estimation of transformation coefficients using multiple years of data when used for the transformation of post-1990s test items.

In contrast to linear conversion, this method converts the test items by restricting the conversion function to a simple additive or subtractive form, which actually restricts the form of the difference between the two test items, and if the form of the difference between the actual test items is not the same, a large error will result. The conversion results of this method are consistent with linear conversion only if the discrimination between the two test items is the same.

This method was first described in Altinok et al. (2018), although it had been used earlier in Hanushek & Kimko (2000).

---

[34]It is also possible to utilize regression, restricting the slope to 1, for direct estimation.

(5) Assumptions on differences in test items: Ratio Linking

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

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

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

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

βIt can be estimated using the same countries that participate in both[35] :

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

where$\mu(Y)$ and$\mu(X)$ are the mean averages of the same countries for the two test items. This method, like linear transformation, allows for the estimation of transformation coefficients using multiple years of data when used for the transformation of test items after the 1990s.

In contrast to linear conversion, this method converts the test items by restricting the conversion function to a simple product, which actually restricts the form of the difference between the two test items, and produces a large error if the form of the actual test item difference is not the same. Only when the difficulty of the two test items is the same does the conversion result of this method agree with the linear conversion.

3. Using the same topic as an anchor: Item Response Theory Linking (IRT Linking)

It is conceivable that if two tests contain exactly the same number of questions, then the scores would be directly comparable even for different groups of students. However, this ideal is not realistic; even the same test varies across rounds. In fact, if we lower the bar and simply have a certain number of overlapping items in the two tests, we can use these repeated items as anchors to achieve comparable scores on the two tests based on a certain methodology.

[35] It is also possible to restrict the intercept term to 0 and use regression for direct estimation.

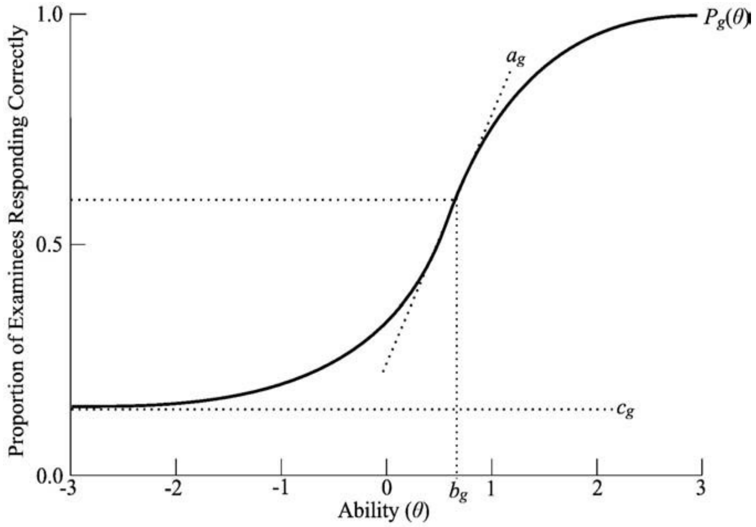
The basis for comparability through identical questions is that both test items rely on Item Response Theory (IRT). Item Response Theory is a modern psychometric theory that has been widely used in international and regional testing programs. Item Response Theory assumes that the probability that a student will correctly answer a given test item is a function of the student's characteristics and the characteristics of the test item.

Specifically, in one of the most commonly used three-parameter logic (3PL) models, for a binary-answer question ($X_{ig} \in \{0,1\}$, where 0 means an incorrect answer and 1 means a correct answer), the Item Response Function (IRF) gives the likelihood of a question being answered correctly by an individual with a literacy ofθ_i :i with a Difficulty ofb_g , a Discrimination of , and a Probability of Guessing Correctly, also known as Guessing. Discrimination) ofa_g and the Probability of Guessing Correctly (also known as Guessing Degree) of$g$ for a question with a literacy ofc_g is given as[36] :

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

whereθ is also known as Latent Variable, usually a wide variety of thinking, abilities, literacies, traits, etc., and in international and regional student tests, usually math literacy, reading literacy, and science literacy. The following figure shows the Item Response Curve (IRC) for the three-parameter model, giving a visual representation of the role of the parameters.

[36] Scores on test items using IRT are largely determined by difficulty and discrimination, which also dictates that conversion connections between tests should be adjusted for difficulty and discrimination.



Note: Cited in Das & Zajonc (2008).

Figure 4-7 Project Response Curve

Depending on the parameters, the above equation can be reduced to a one-parameter model (difficulty) and a two-parameter model (difficulty, differentiation). With the knowledge of whether the students answered correctly or not and the characteristic parameters of the question items, the student characteristic parameters (i.e., student's ability and literacy) are estimated based on the model to obtain the student's test item scores. Item response theory has two key assumptions: the assumption of unidimensionality of the latent variable being measured and the assumption of parameter invariance. The assumption of unidimensionality means that all items that make up a given test measure the same latent trait; the assumption of parameter invariance means that the topic characteristic parameters are fixed for any population and are not affected by the distribution of examinee ability.

In principle, if two test items have a certain number of identical questions, it is possible to connect two test items based on the same questions and realize the comparability of the scores of the two test items. Suppose, for example, we take the test program that has been implemented for three rounds and each round of test has only two questions, the questions of each round are (Question 1, Question 2) (Question 2, Question 3) (Question 3, Question 4), and the two rounds in turn have the same question items: (Question 2) and (Question 3). When the characteristic parameters of question 1 are fixed and known, the characteristic parameters of question 2 can be determined since both question 1 and question 2 should estimate the same ability (latent variable). Similarly given the characteristic parameters of question 2, the characteristic parameters of question 3 can be estimated. By analogy, it is possible to correlate the third and second round test items with the first round test items to achieve comparable test items across multiple rounds (Das & Zajonc, 2010). It is important to note that although the previous example was given with one duplicate question item, in practice, the two test items should have a certain percentage of duplicate items to achieve the goal of reducing linkage error (Hastedt & Desa, 2015).

Indeed, after the 1990s, international and regional test programs have included a certain amount of repeated question items in different rounds of test items as a way to achieve direct comparability of test item scores across rounds in time (Angrist et al., 2021) [37] .

Depending on the feature parameter settings of the question items, different approaches arise, and we focus here on the following two ways of constructing connections (Sandefur, 2018):

(1) Equivalence of the characteristic parameters of the question term

In this case, the same question items from the target test item (typically a regional test item, such as SACMEQ) are utilized, using the parameters that characterize these same items from the reference test item (typically an international test item, such as TIMSS)[38] .

That is to say:

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

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

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

The subscripts$r$ and$t$ represent Reference Tests and Target Tests respectively.

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

[37] See Hanushek & Woessmann (2012a) for studies that have converted pre-2000 test scores to tests that are comparable across time, typically the U.S. NAEP tests.

[38] The validity of the method relies on the Differential Item Functioning (DIF) test.

(2) Equivalence of question difficulty and differentiation (Mean-sigma Linking)

Unlike item characteristic parameter equivalence, instead of aligning the characteristic parameters of the same items on the target and reference test items, it requires ensuring that the average difficulty and discrimination of the same items remain constant across the two tests[39] . Based on the invariance assumption in the core assumptions of item response theory, both equivalent test item scores can be related by a linear transformation:

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

Consistent with the previous ones in the same aggregate (country), coefficient A is used to adjust for discrimination and intercept B is used to adjust for difficulty. Similar transformations can be applied to the characteristic parameters of the question items:

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

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

Theoretically, it would be possible to obtain bothA_{rt} andB_{rt} parameters with only one identical question, however, measurement error and imperfect fitting of the model made this route practically unworkable. In fact, the researcher is forced to choose between the different$A$ and$B$ obtained for each question.

Alternatively, a simpler approach, called the Mean-sigma method, is taken to obtain$A$ and$B$ from the mean and standard deviation of different eigenparameters$b$ for the same question items:

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

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

After obtaining the conversion coefficients, i.e., they can be directly applied to the students' literacy parametersθ_i , converting the target test item scores to the reference target scores. That is, this approach can be viewed as estimating the parameters of the linear conversion formula in the same aggregate (country) from the level of the question

[39] There is no mention in the literature of how the guessing degree is set.

item characteristic parameters. Literature using this approach includes Sandefur (2018).

(iii) Literature summary and presentation of a database of globally comparable student cognitive skills

1. Summary of literature

We give the literature that employs the basic methodology presented in the article.

Table 4-3 Different methods and literature for correlating different test items

<table>
<tr><th>form</th><th>suppose that...</th><th colspan="2">Methods and Literature</th></tr>
<tr><td>an identical individual</td><td></td><td colspan="2">Patel & Sandefur (2020)</td></tr>
<tr><td rowspan="8">Same overall (country)</td><td></td><td>Identical totals (same overall): merge test information</td><td>Same countries (with duplicates): expanding the number of countries in the sample</td></tr>
<tr><td rowspan="4">Similarity based on microdistribution</td><td>Mean standard deviation conversion
Reardon et al. (2021)</td><td></td></tr>
<tr><td rowspan="3"></td><td>Mean standard deviation conversion
Hanushek & Woessmann (2015); Gust et al. (2024)</td></tr>
<tr><td>isozyme conversion
Sandefur (2018)</td></tr>
<tr><td>Mean standard deviation conversion applied to macro data
Hanushek & Kimko (2000), Hanushek & Woessmann (2012a), Hanushek & Woessmann (2012b); Altinok et al. (2018)</td></tr>
<tr><td rowspan="2">Based on test item differences</td><td rowspan="2"></td><td>linear transformation
Angrist et al. (2021)
mean value connection
Hanushek & Woessmann (2012a)</td></tr>
<tr><td>ratiometric connectivity
Altinok & Murseli (2007); Altinok et al. (2014); Altinok et al. (2018); Patrinos & Angrist (2018); Altinok and Diebolt (2023)</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>same topic</td><td colspan="3">Equivalence of itemized feature parameters</td></tr>
</table>

	Das & Zajonc (2010), Singh (2014), Sandefur (2018)
Equivalence of question difficulty and differentiation	Sandefur (2018)

2. Introduction to the database

Currently, there are a number of studies devoted to the work of merging numerous international and regional student test score data ï¿½ yet only a few of these studies have publicized the databases they have developed, and even fewer have promised to update them on an ongoing basis. We summarize the known internationally comparable databases on students' cognitive skills in Table 4-4.

The first is the HLO database provided by the World Bank, which is currently the only one that is being continuously updated. The latest version of the HLO data, constructed by Angrist et al. (2021), was built using mainly linear transformation methods, and the database can be searched directly in the World Bank's data bank. As with previous versions of the HLO database, the HLO database has been used in several World Bank reports and is included in the World Bank's Human Development Index Database (HCI Database).The HLO database's wide range of meta-source test items allowed it to provide data for 164 countries and districts for the years 2000-2017, by gender-stage of education (primary, middle school) - National mean data by subject (math, science, reading), however, due to test item limitations, this data is unbalanced panel data, and many countries have relatively high levels of missing student test scores from earlier years.

Next is the data assembled by Lim et al. (2018), which first uses a linear transformation method to construct comparable student cognitive skills data, and then utilizes spatiotemporal Gaussian process regression to estimate the learning index for different country-gender-age groups (Learning index.) This data provides balanced panel data for 186 countries from 1990 to 2016. However, in reality, only 132 countries have data from the test program and not all years are covered, and the results for most countries and time periods in the data are generated by fitting, the reliability and accuracy of which is not yet known, and which needs to be noted when using the data. .

Then there is the Learning Skills data constructed by Altinok and Diebolt (2023), which was designed to construct quality-adjusted data on years of schooling (see next subsection), but also gives data on students' cognitive skills in the Supplementary Material. This data is largely based on ratio-linkage and synoptic linkage transformation methods, and then the same value-added procedure is utilized to add values for years and countries with missing data. Note that while it indicates in the article that its students' cognitive skills are broken down by gender-stage of education (elementary, middle school)-subject (math, science, reading), it gives the final data as a single value[40] .

Finally, the data assembled by Gust et al. (2024), constructed primarily using mean-standard deviation transformations, is the most recent result from the Hanusheck team. The data is appended to the accompanying article. Since Gust et al. (2024) mainly wanted to construct the most recent data on students' cognitive skills in each country, most of the student test items used in time are from 2018 or 2019, so this data is a bit more up-to-date in time than the HLO database, and only provides single-period data on students' test scores in each country, which is a cross-sectional data; in terms of the student population, it is mainly based on junior high schools, and there are no middle school countries, supplemented with elementary school ones; in terms of subjects, it provides the average of math and science scores, but not sub-subject scores; and in terms of the number of countries, it covers 159 countries and territories, which is also the smallest among all the data.

[40] No description of this value was found, but it is suspected to be the average value for all cases.

Table 4-4 Publicly Available Student Cognitive Skills Database

	HLO database (new version)	Global Human Capital Estimates	Learning skills	Student Access Data on a Uniform Scale
Methodological literature	Angrist et al. (2021)	Lim et al. (2018)	Altinok & Diebolt (2023)	Gust et al. (2024)
Conversion method	Linear conversion is the main focus, with mean-standard deviation conversion as a secondary focus	Linear transformation+ Fitting complementary values	Ratio-linkage-based and iso-linkage conversion with complementary values	Primarily mean-standard deviation transformations, supplemented by high-fit complements
Metadata sources	timss, pirls, pisa, sacmeq, pasec, serce, terce, egra	PISA, PIRLS, TIMSS, IAEP, Reading Literacy Study (1991), SACMEQ, LLECE, PASEC, NAEP, NAS, IQ, MYS	timss, pirls, pisa, sacmeq, pasec, serce, terce, egra, aser	timss, pisa, terce, serce, sacmeq, pasec
Whether the data are sex-disaggregated	be	clogged	indiscriminate	clogged
segments	Primary+ Middle School	5-9, 10-14 and 15-19	indiscriminate	Primarily middle school
subjects	Mathematics (Math), Science (Science), Reading (Reading)	Learning index	indiscriminate	Math and Science Means
Data span and structure	Unbalanced panels, 2000-2017	1970-2016 Balanced panels	Balanced panels 1970-2020 (part of the time for panels)	2018/2019 cross-section
Number of countries and territories	164	186	167	159
Data Acquisition	https://datacatalog.worldbank.org/search/dataset/0038001	https://ghdx.healthdata.org/record/ihme-data/global-human-	https://link.springer.com/article/10.1007/s11698-023-00276-x	Gust et al. (2024) Schedule of Articles

V. Integrating years of schooling and students' cognitive skills: a quantitative and qualitative approach

(i) Constructing both qualitative and quantitative indicators

In the process of moving from a focus on the quantity of educational acquisition to a focus on the quality of educational acquisition, the academic community has not completely rejected the importance of the quantity of education and discarded the average number of years of schooling, but has considered that the quantity and quality of educational acquisition are equally important. Some studies have therefore attempted to combine quality and quantity of education, and two types of pathways have emerged: linear and multiplicative.

Before entering the analysis, we need to return to the question that the average number of years of schooling can only measure the quantity of educational attainment, which has become a consensus through the efforts of numerous scholarly studies. However, do students' cognitive skills measure the quality of educational acquisition?

According to the theory of human capital and the theory of educational production function, cognitive skills are an important part of human capital, which is the common output of input factors such as school education, family education, and individual talent, etc. For school education, existing studies in turn usually decompose it into two simple aspects: the quantity of education and the quality of education. That is to say, from a doctrinal point of view, years of schooling points to the quantity of schooling received by an individual, which is a kind of education input indicator; while students' cognitive skills is an education output indicator, which is the joint output of education quantity and education quality, rather than a direct measure of education quality.

So it seems that years of education and students' cognitive skills are two indicators of different dimensions and its not possible to simply blend them together.

Based on the above discussion, the correct approach should be to separate the quality of education from student cognitive skills using the education production function. At the same time, when students are in the same grade or age (i.e., the same amount of education), student cognitive skills are only determined by the quality of education, and student cognitive skills can be used to reflect the quality of education. That is, we can use student cognitive skills to measure educational quality only when students are in the same grade or age; when students do not exist in the same grade or age, it is necessary to use the educational production function to estimate educational quality from student cognitive skills. [41]We will return to this point in Chapter VIII.

The above analysis also shows the reader that the key to the use of student cognitive skills is not only the comparability of test items, but also the consistency of students. However, existing research on the use of student cognitive skills has typically prioritized "data availability over student agreement": that is, filling in the gaps with data from other tests, disregarding or forcibly ignoring student age and grade level inconsistencies (Gust et al., 2024).

There have also been studies that estimate the quality of education received by adults at different times based on adult cognitive skills and calculate quality-adjusted years of schooling in combination with the number of years of schooling an adult has had (Hanushek & Zhang, 2009). Unlike student cognitive skills, the use of adult cognitive skills necessitates the use of an educational production function to estimate the quality of education from adult cognitive skills.

In the following we provide a detailed description of this aspect in terms of the human capital (cognitive skills) production function.

It is important to note that most of the following talk assumes that school quality cannot change over time and grade levels, i.e., the quality of education this year is the same as tomorrow's; the quality of education in grade 3 is the same as in grade 4.

(ii) Human capital (cognitive skills) production functions and linear and product combinations

1. Human capital (cognitive skills) production function

According to the theory of educational production function, educational output is the final product of schooling, home education, and out-of-school education (Hanushek & Zhang, 2009). Among these factors, the effect of schooling on cognitive skills is a function about the quality and quantity of education. Combined with Hanushek & Woessmann (2012a), the human capital (cognitive skills) production function can be

[41] When using cognitive skills from the same grade level as the quality of education, the study does not create a problem because the amount of education at this point is the same; when using cognitive skills from different grades as the quality of education, there is some error in using students' cognitive skills as the quality of education at this point because the amount of education at this point does not want to be the same.

expressed in the following equation:

$$H(CS) = \lambda F + f(n, q) + \eta A + \alpha Z + \nu \quad (5\text{-}1)^{42}$$

where H is human capital, in this case primarily student/adult cognitive skills (CS); F is household inputs, $\phi(n, q)$ is school inputs, including the quantity (n) and quality (q) of schooling[43], A is individual talent, and Z is other factors, including labor market experience, health, and so on.

Just as "Schooling is not learning", "Learning is not just schooling", however, research often focuses more on schooling. Because schooling has a greater impact on cognitive skills than out-of-school education (Filmer et al., 2020), other components are often overlooked by some studies:

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

The most important message of this production function is that schooling does not necessarily lead to substantial development of individual skills and effective accumulation of human capital. The quantity and quality of schooling are only inputs to the development of individual skills, and the transition from schooling inputs to individual skill development is a function of $f(Schooling, Quality)$.

However, the functional form of $f(Schooling, Quality)$ is unknown and needs to be set artificially. The discussion about the functional form can be divided into two points: the way in which the quantity and quality of education are combined, but rather the efficiency of educational production, i.e. how the combination of quantity and quality of education affects educational output. Currently, there are two forms of combining quantity and quality of education: linear and product.

(1) Linear binding

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

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

In literature such as Hanushek et al. (2017), Angrist et al. (2020) and others, the

[42] For the sake of simplicity, the notation of discipline and time is ignored in the above.
[43] This refers to unit mass.

exponential function form is specifically used:

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

where s is the quantity of education, usually using average years of schooling; Q is the quality of education, usually using students' cognitive skills; and$r$ and$w$ are usually derived from micro-account income equation estimates.

Firstly it is easy to see that after taking logarithms on both sides, this form is very close to the account income equation, which is probably why the logarithmic form was taken. Secondly, existing studies often use the cognitive skills of students of different grades or ages directly as the quality of education put into the regression will, ignoring the fact that the cognitive skills of students of different ages or you there are not exactly 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 the form of a product in a function:

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

Unlike before, in order to make this product meaningful,Q is usually the quality adjustment factor here, rather than the direct quality of education, so thatSQ can be interpreted as quality-adjusted years of schooling[44] . Formally, the formula can be expressed as follows:

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

Where $QAYS_c$ is the adjusted years of schooling, S_c is the average years of schooling in the country c and Q_c^b is the quality adjustment factor. The quality adjustment factor, which is generally based on a country, can be calculated using the following formula:

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

Whereq_c is the quality of education in the target country andq_b is the quality of

[44] These studies include Hanushek & Zhang (2009), Kaarsen (2014), Filmer et al. (2020),, Reiter et al. (2020) , Glawe & Wagner (2022) and others. In some studies, quality-adjusted years of schooling are directly equivalent to human capital, and in others, its still a function to which a layer is applied to calculate human capital, e.g. Kaarsen (2014). The confusion in the use of related concepts can also be seen here.

education in the Benchmark Country.

$QAYS$ In effect, the average years of schooling are transformed to include information on the quality dimension without discarding the average years of schooling.

2. Choice of function form: different types of cognitive skill production functions

For linear combinations, the crux of the problem lies in estimating the coefficients r and w for the combination of both quantity and quality, which, as noted earlier, are largely estimated from microdata.

For multiplicative combinations, on the other hand, the crux of the matter is to estimate the quality of education. The quality of education reflects a schooling system's productivity (Productivity; Filmer et al., 2020) or schooling system effectiveness (Effectiveness; Kaarsen, 2014).

Estimates of education quality also need to go back to the human capital (cognitive skills) production function.

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

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

CS_c is the cognitive skills of students/adults for country c, $n_{c,l}$ is the number of years of education corresponding to the grade l for country c, and in existing international tests of students, l is usually the 4th or 8th grade, and q_c is the quality of education received by students in different grades of the education system for country c. Since it is assumed that the quality of education does not vary by grade, the symbol for quality of education is subscripted only by the country c[45].

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

[45] This can also be interpreted as an estimate of the average quality of the education system.

perspective:

One is the idea that the quality of education is an input and that there is a question of productivity between inputs and outputs. Inputs can have diminishing marginal returns, as in Kaarsen (2014),$f(\cdot)$ in logarithmic form:

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

Inputs can also have constant marginal payoffs, 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 the idea that the quality of education is an output, and that the accumulation between outputs equals the total output. In the context of this paper, this means that the quality of education is the cognitive skill score that can be obtained each year. The cognitive skill production function can then be simplified directly to[46] :

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

The quality of education can be obtained by selecting the appropriate functional form and imposing specific assumptions from different perspectives.

Once the quality of education has been estimated using the cognitive skills production function, the formula can be used to construct a quality adjustment factor to calculate the adjusted average years of schooling.

3. Calculation of QAYS: general steps are points of divergence

Thus the general sequence of such studies is to estimate the quality of education q for each country for the corresponding period using data on students' cognitive skills and equations (26) or (27); then, based on the quality of education obtained for each country, to calculate the quality of education adjustment factor$Quality^{benchmark}$ using equation (29); and finally to obtain the QAYS data using data on the average years of schooling for each country for the corresponding period, taking advantage of

[46]In this scenario, if the quality of education can vary with different grade levels, the production function can be written in cumulative form:

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

equation (28).

In our calculations, we used the number of students' cognitive skills at a given point in time, such as the scores of 15-year-old students in PISA 2018, to estimate the quality of education from the student cognitive skills data[47] . As well as data on average years of schooling at a given point in time, the choice of time point for the adult data varies from using data on the recent graduates close to the time of the students, such as the average years of schooling for the 25-29 year olds as of 2018 (Filmer et al., 2020), to data on the average years of education for the 25-64 year olds for the corresponding time period (Kaarsen. 2014; Altinok & Diebolt, 2023). The choice of the population group of adults with average years of schooling determines whether QAYS measures a country's current education system or its stock of human capital. Regardless of the choice, however, student data and adult data are not contemporaneous. Theoretically, however, the data for adults aged 25-29 are closer in time to the student data and are more robust, so the most appropriate use of QAYS is to measure the quality of the current education system.

In other words, the calculation of QAYS presupposes the availability of data at two points in time: data on students' cognitive skills at one point in time and data on the average number of years of adult schooling at one point in time. Of these, student cognitive skills are used to estimate the quality of education, which is then adjusted for average adult years of schooling.

From the above derivation process, it can be seen that the QAYS indicator is derived by applying layers of assumptions on the way of combining the quantity and quality of education, as well as the form of the human capital production function, which means that the implementation of the study relies on the assumptions on the cognitive skills production function, and at present, the academic community still has insufficient knowledge of the techniques and ways of human cognitive skills production, which makes scholars conduct This has forced scholars to make many subjective assumptions that are not supported by much empirical evidence, so the construction of the QAYS index has a greater risk of measurement bias. The "black box" of cognitive

[47] Ideally, we should use the quantity and quality of education at the same point in time for integration. Theoretically, we should use the quantity of education at age 15, but this is a red herring; after all, students' cognitive skills are already an indicator of both quantity and quality of education.

skill production function is an insurmountable mountain in front of scholars in this field.

(iii) Literature summary and database presentation

1. Literature summary

The following table gives the literature using the above basic approach.

Table 5-1 Angle-Function-Documentation Correspondence Table

input and output	function form	literatures
input side	Logarithm ($\beta * ln(nq)$) Linear ($\beta * nq$)	Kaarsen (2014) Hanushek & Zhang (2009)
output side	Product (nq)	Filmer et al. (2020) Angrist et al. (2020) Glawe & Wagner (2022) Altinok & Diebolt (2023)[48]

2. Introduction to the database

Currently, there are two international databases of QAYS available, one developed and constructed by Filmer et al. (2020) and the other constructed by Altinok & Diebolt (2023). Since both databases use student cognitive skills in the construction of , and student cognitive skills are "Learning", both databases also refer to quality-adjusted years of schooling as "Learning-adjusted Years of Schooling" (LALS), which is a term used to describe the number of years of schooling. adjusted Years of Schooling (LAYS)".

The LAYS data developed by Filmer et al. (2020) has been adopted by the World Bank and it is also included in the HCI database for the calculation of the Human Capital Index (HCI). The World Bank HCI database provides unbalanced panel data for 174 countries for four years, 2010, 2017, 2018, and 2020. It is important to note that Filmer et al. (2020) utilizes years of schooling data for the 25-29 year old population, close to the year of their student cognitive skills data, so the data measures a country's current education system[49] .

The other is constructed by Glawe &Wagner (2022), which differs from Filmer et al. (2020) mainly in that it constructs equilibrium panel data for 33 countries for the

[48] The article provides fewer technical details and contains contradictory points that the reader should be aware of.

[49] In a subsequent update of the data, some details were not presented in the explanatory document, so it is not yet known how it was possible to reach a coverage of 174 countries.

period 1995-2015 (5-year interval).

The third data is constructed by Altinok & Diebolt (2023), which provides balanced panel data for 120 countries from 1970 to 2020 (5-year intervals). Altinok & Diebolt (2023) differs from the previous two in two ways. First, it uses years of schooling for the population aged 25-65, so the data points to a country's human capital stock rather than its current education system. For two, the data makes more fitted estimates, such as the average years of schooling in each country in 2020, the year and the cognitive skills of the students who are partially missing in the country, and as can be seen here, its data is not pure and may have a large margin of error. As it is a newly constructed database, thus the extent to which the data is applied and accepted is not yet known.

Table 5-2 Database on quality of education or adjusted mean years of schooling

		LAYS database		
Methodological literature		Filmer et al. (2020)	Glawe &Wagner (2022)	Altinok & Diebolt (2023)
Methods (in function form)		Formulas (5-11)	Formulas (5-11)	Formulas (5-11)
Metadata sources	Student Cognitive Skills	In the article: TIMSS or PISA Follow-up Update: Global Dataset on Education Quality (2020 Update)	TIMSS or PISA	timss, pirls, pisa, sacmeq, pasec, llece, serce, terce, egra, aser
	educational attainment	Barro-Lee data (25-29 years old)	Barro-Lee Data+ CSL Database[50] (25-29 years old)	Barro-Lee data (25-64 years)
time span		In the article: 2015 Subsequent update: unbalanced panels, 2010-2020	1995-2015 balanced panels (5-year intervals)	Balanced panels 1970-2020 (5-year intervals - part of the time for panels)
Number of countries and areas covered		Article: Follow-up update: 174	33	120
Data Acquisition		https://datacatalog.worldbank.org/search/dataset/0038030/Human-Capital-Index	uncharted	https://link.springer.com/article/10.1007/s11698-023-00276-x#Sec110

VI. Adult cognitive skills: an optimal measure of human capital

The discussion in this paper moves from average years of schooling, to students' cognitive skills, to a combination of quantity and quality of education. One cannot help but ask the question, are all of the above measurements the best measure of human

[50] https://www.parisschoolofeconomics.eu/en/cohendaniel/international-educational-attainment-database/

capital? If not, which indicator would it be?

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

There are several reasons for this: first, student cognitive skills are a flow indicator, whereas adult cognitive skills data are an indicator of human capital stock; this is partly because participants in adult cognitive skills tests cover all ages in the labor market, whereas student cognitive skills are usually only available for a particular grade or age, and partly because student cognitive skills data are available for a particular grade or age, not for all the years of their cognitive skills after education.

Second, according to the education production function, both the quantity of education (average years of schooling) and the quality of education are inputs, with the cognitive skills of adults being the final output. Third, human capital is the elements of knowledge and skills that are condensed in a person, so it is also clear that a measure of human capital synthesized with the quantity and quality of education (QAYS) is not as good as a direct measure of human capital.

(i) Constructing a database of adult skills: using linkages with student data to break through the limitations of the adult survey to include a limited number of countries

Existing surveys of adult cognitive skills[51] include the International Adult Literacy Survey (IALS)[52] , the International Adult Literacy and Life Skills Survey (ALLS), the Program for the International Assessment of Adult Competencies (PIAAC) , and the Skills Towards Employability and Productivity Survey (The Skills Towards Employability and Productivity). [53]The International Adult Literacy Survey (IALS)[54] , the International Adult Literacy and Life Skills Survey (ALLS), the Program for the International Assessment of Adult Competencies (PIAAC) , and the Skills Towards

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

[52] The survey was conducted by the OECD between 1994 and 1998 in 22 countries and covered prose literacy, document literacy and quantitative literacy.

[53] The survey, implemented by the OECD between 2003-2007 and treated as a successor to the IALS, surveyed 11 countries, replacing the previous Quantitative Literacy with Numeracy in terms of content, and adding Problem-solving.

[54] The survey was implemented by the OECD between 2011 and 2018, surveying 37 countries, and in terms of content, it consists of Literacy, Numeracy, and Problem-solving in Technology Rich Environments. The program is designed to be used in a variety of countries, including the United States of America.

Employability and Productivity (STEP) program. Productivity (STEP) program[55] . Of these, PIAAC covers the largest number of countries, but data are available for only 36 countries.

It can be argued that the limited number of countries participating in PIAAC greatly limits its use in macro studies. In contrast, with the student survey test program, more and more countries are included. For this reason, a few scholars have recently begun to attempt to link student cognitive skills and adult cognitive skills data across countries, and to expand the sample size of countries with adult cognitive skills data by constructing a transformational relationship between the two (Égert et al., 2024).

(ii) Birth Cohort Matching and Association Functions for Students and Adults

In the research direction of expanding data on adult cognitive skills, the authors are limited by their vision and have only found one study so far: Égert et al. (2024). Therefore, this section focuses on the methodological ideas of this study.

The idea behind this approach is similar to the relationship between flows and stocks. The educated student, as a flow, continuously enters the labor market and becomes a part of the adult population, while the adult population, as a stock, is continuously replaced by the flow under the continuous integration of the flow. Thus, today's adult labor stock can be viewed as the result of the constant substitution of students from previous periods. Therefore, if one has data on all periods of flows that have replaced all adults (i.e., the cognitive skills of students in all previous historical periods), one can estimate today's stock (i.e., the current cognitive skills of adults) using the flows of the past.

As shown in the figure, suppose that 15-year-old students in a country c participated in the 1995-2015 PISA test, which was scored as a flow, and adults in that country participated in the 2017 PIAAC test, which was scored as a stock. If the cognitive skills of the 15-year-old students who participated in the 1995-2015 PISA tests do not change again after their participation in the tests, we can construct data on the cognitive skills of the country's adults aged 15-39 years old in 2017 based on the

[55] The survey was conducted by the World Bank (World Bank) in 17 non-OECD low- and middle-income countries between 2012-2017, and unlike the surveys implemented by the OECD, it focuses on Reading Skill.

birth cohort corresponding to the 15-year-old students who participated in the 1995-2015 PISA tests in 2017$PISA'$ 2017. In other words. Adult cognitive skills of 15-19 year olds in the figure$PISA'$ 2017 are equal to the cognitive skills of 15 year old students who participated in PISA 2015, those of 20-24 year olds are equal to the cognitive skills of 15 year old students who participated in PISA 2015, and so on (as shown by the solid line with arrows in the figure).

However, the fact is that the$PISA'$ 2017 we constructed above will not really equal the adult cognitive skills in Country C in 2017 due to postsecondary education, on-the-job training, and skill depreciation. In order to solve this problem, the relationship between the constructed adult cognitive skills ($PISA'$ 2017) and the true cognitive skills (PIAAC 2017) needs to be known. By using the corresponding data for each birth cohort in the constructed$PISA'$ 2017 and the real PIAAC2017 (as shown by the dotted line in the figure), we can then estimate the linking transformation equation, based on which the purpose of using the PISA scores of students' cognitive skills in the previous period to derive the present-day adult cognitive skills, the PIAAC, can be realized.

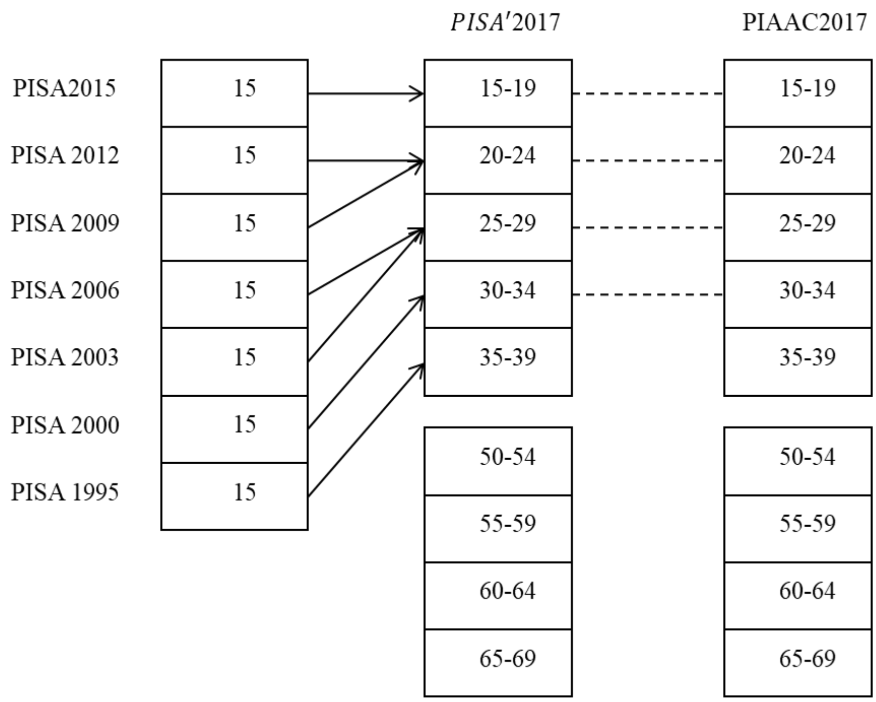


Figure 6-1 Simple Schematic of the Research Idea

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

In the first step, countries that participated in both the PISA and PIAAC tests were selected as anchor countries. Data on adult cognitive skills were extrapolated/fictionalized for the corresponding period according to when students took the tests, as shown above at$PISA'$ 2017。

In the second step, the conversion function is constructed using the fictitious adult cognitive skills data of the above countries (e.g.,$PISA'$ 2017) and their real adult cognitive skills (e.g., PIAAC 2017) with birth cohort as the data unit. This is the most important step, and the quality of the conversion function is directly related to the quality of the converted data afterwards.

In the third step, the constructed transformation function is applied to analyze the data for countries that have participated in many periods of PISA testing to estimate the cognitive skill scores for adults in each birth cohort in those countries. This is then done for all birth cohort scores in that country, and the mean or other statistic for overall adult cognitive skills in that country is obtained. Since the estimated adult cognitive skills data are for the year 2020, countries that participated in the PIAAC test also participated in the estimation of .

Égert et al. (2024) utilize the pioneering use of historical annual human capital flows to cumulatively calculate the current stock of adult human capital. However, the method has its limitations; it is more applicable to high-income economies with high basic education enrollment and transition rates, and for economically and educationally less-developed countries and regions, where a large proportion of the population does not have access to basic education and where the PISA scores of students in these countries are not representative of the entire population of the birth cohort to which they correspond, the use of the method may lead to a biased measurement of the cognitive skills of adults in developing countries. biased measurement.

Using the methodology described above, Égert et al. (2024) constructed human capital stock databases in 2020 for the population aged 15-64 in 17 countries and for the population aged 15-39 in 54 countries, with the population aged 15-39 covering a sample of 18 more countries than PIAAC. Since in practice the implementation of the methodology is heavily dependent on whether or not students in basic education in the sample countries have been tested internationally in the past and for how many rounds, and since high-income economies are the "usual suspects" in international student testing, the sample constructed by Égert et al. (2024) is still dominated by high-income economies and includes only a small number of middle-income economies. Therefore, Égert et al. (2024) constructs a sample that is still dominated by high-income economies, with only a small number of middle-income economies, which account for 74.07% of the countries covered by the data on the population aged 15-39, and 68.52% of the OECD countries.

(iii) Literature summary and introduction to the adult cognitive skills database

1. Summary of literature

As mentioned earlier, there is only one article in the current literature.

Table 6-1 Summary of Literature

methodologies	literatures
Birth queue matching and correlation	Égert et al. (2024)

2. Introduction to the database

The paper Égert et al. (2024) constructs human capital stock data for the population aged 15-64 for 17 countries in 2020 for the population aged 15-39 for 54 countries, where the data for the population aged 15-39 covers a sample of 18 more countries than PIAAC. Due to the limitations of the data on students' cognitive skills, the human capital stock indicators it constructs are still more for developed economies and include only a smaller number of middle-income economies.

In addition, the data constructed for this paper are not publicly available and need to be obtained from the authors themselves[56] .

[56] Tried it, no reply.

VII. Labor market wage information: a direct reflection of education quality and human capital

(i) Research intention: to isolate information on the quality of education and human capital across countries from information on wages

Theoretically, the quality of education will be directly reflected in the performance of graduates, and the wages of the labor force in the market can be used as a direct reflection of an individual's human capital (Lee & Barro, 2001). Therefore information on the wages of the labor force in the market can be used to reflect the quality of education and human capital in each country. There are two ways of obtaining these two types of information from wage information: one is based on the Accounts Income Equation, which uses the rate of return to education as a measure of the quality of education in each country. The other is based on the macro-production function, which isolates the effect of human capital from wage income.

Table 7-1 Two Ways to Use Payroll Information

infrastructural	Bias of separated measurement indicators
Based on the Account Income Equation	Quality of education
Based on the macro production function	human capital

One problem with these two approaches is that the wage performance of the labor force across countries depends not only on the schooling and human capital of each country, but also on the external environment (Lee & Barro, 2001); and the inconsistency of labor market environments across countries makes wage information not comparable across countries. Gaps in labor earnings do not only reflect differences in the human capital of the labor force, but may also reflect differences in the labor market, such as total factor productivity, the level of job information availability, and so on. Therefore, existing studies usually use data on immigrants from other countries in one country's labor market for sample selection, which ensures labor market consistency.

Two issues must be considered when using migration data: first, migration may be selective, making migrants from a particular country not representative of that country as a whole. This selectivity comes from two main sources: the self-selection of immigrants, whereby more educated people from low-income countries are likely to migrate to countries with a richer level of economic development, and therefore using a sample with higher individual ability may overestimate the quality of education in that country; and the selection of the country of entry, whereby countries such as the United States and the United Kingdom usually set conditions to screen immigrants. Second, there is the question of whether skills are fully transferable, i.e. whether skills learned in the home country become irrelevant after migration. If this is the case, the estimated market rate of return does not represent the true quality of the education system in that country.

(ii) Treatment: the Account Income Equation under the micro-path and the price of human capital under the macro-path

1. Account income equation

When using a country's immigration data to construct the account's income equation, it has the following formula:

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

where$c2$ refers to the country of immigration,$c1$ refers to the country of origin of the immigrant;$log(W)$ is the logarithm of the individual's income,S is the individual's years of schooling,μ is the estimated coefficient of interest, which represents the quality of education in the country of origin of the immigrant and indicates the income boost that can be generated by one year of schooling; and$X$ is other control variables.

Using data on immigration in a given country (e.g., the United States), it is possible to estimate the quality of education in each countryμ^{c} .

Currently, there is limited research using this method to isolate the quality of education, and in the case of the two issues of immigrant selection and skill transferability of immigrant data, the existing research has only verified that the issues of immigrant selection and skill transferability do not have a large impact on the estimation results (Schoellman, 2012).

2. Price of human capital

This approach starts from a macro production function. The standard production function is:

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

where Y_c is gross output; K_c^{α} is physical capital stock; A_c is total factor productivity;$H_c = h_c L_c$ is total labor input,h_c is human capital per capita, andL_c is the number of workers.

For production, the following equation needs to be maximized in order to maximize the benefit:

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

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

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

It can be seen that the price of skills in each country (w_c) is influenced by the total factor productivity of each country as well as the capital-output ratio, among other things. In addition, if workers are paid at marginal output, then workers earn:

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

From there.

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

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

where$\log(h_c) = E_c[\log(h_{i,c})]$. From the above equation, it is clear that the human capital of each country can be obtained through fixed effects if we know thez_c of each country.

Therefore, usually this type of study is usually divided into two steps, the first step is to first estimate thez_c ; the second step is to use the above equation to obtain the human capital of each country.

The following is specified in the simplest case: when there is no immigration choice or skill transferability problem, the human capital of immigrants does not change with immigration, and therefore:

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

By estimating the above equation, it is possible to obtainz_c for each country, although this estimation is extremely data-demanding, requiring pre- and post-migration wage data and the availability of multiple immigrant-source and immigrant-immigrant-immigrant countries[57] .

The problem is complicated when there is an issue of immigrant choice and skill transferability, which some studies have addressed by internalizing immigrant choice and skill transferability (Martellini et al., 2024).

(iii) Introduction to the database

1. Summary of literature

The literature in this section is summarized below:

Table 7-2 Summary of Literature

methodologies	literatures
The Account Income Equation under the Micro Pathway	Schoellman (2012)
The price of human capital in the macro path	Martellini et al. (2024)

2. Introduction to the database

Neither of the two studies of interest in this paper made the self-constructed data publicly available as results, and only Martellini et al. (2024) provided reproduced data for the article's graphs in a data annex, and thus cannot be expanded upon here.

VIII. Framework for analyzing human capital indicators

We have presented the history of the evolution of human capital indicators and the five key measurements indicators along the evolutionary journey, however, existing analyses of the indicators are scattered and not put into the same framework.

[57] Note that when all immigrants come from a single country, or when data on immigrants from only one country are available, the equation can only estimate results as the difference between the price of skills in each country and the country of immigrants, but this difference does not affect the human capital estimates for each country.

(i) Analytical framework for human capital indicators in the harmonized framework

Starting from the human capital and education production functions. We begin by placing the direct-indirect, stock-flow, output-input, quality-quantity, labor market-education system into the same framework.

(Direct - Indirect) Firstly, direct and indirect. According to the theory of human capital, human capital is the sum of qualitative factors such as knowledge, skills and physical strength (health) that exist in the human body and have economic value. Therefore, if there is such a thing as an optimal measurement of human capital, it must refer directly to the knowledge and skills of the human being, and any measure that does not involve such elements as knowledge and skills, health and so on, is only an indirect measure of human capital.

(flows-stock/labor market-education system) Second, when we talk about a country's human capital, we usually point to the human capital of all adults in the labor market. Therefore, the measurement of a country's human capital should be the measurement of the stock of a country's human capital, whereas groups that have not yet entered the labor market, such as students in the education system, are the source of the stock of human capital, which is an indicator of flows that have not yet occurred.

(Output-input) Cognitive skills are an important component of skills. According to the education production function, **the cognitive skills of a given birth cohort in a given period can be considered as a function of the cognitive skills acquired by adults of that age during their school years; (quantity-quality) the cognitive skills acquired during their school years can be considered as a function of the minimum number of years of schooling required to complete all levels of education in the country (i.e. the quantity of education) and the quality of education received (the quality of education can be considered as the average quality of education received); and the quantity and quality of education are in turn a function of the quality of education received in a country over time. The quantity of education is a function of the quality of education received (the quality of education can be seen as the average quality of education received); and the quantity and quality of education are in turn a product of the long-term development of a country's education, the result of long-term investment in that country's education, and a function of the relevant inputs to the various levels and types of education (e.g.,**

teacher-student ratios, per capita expenditures). The same factors (quantity and quality of education) can be both inputs and outputs in different contexts. **If we consider the production of education as a dynamic process of transformation of inputs and outputs at different levels, at the final stage, the end product of a country's investment in human capital should be the cognitive skills of the adult population, with the cognitive skills acquired during the student years as its input; at the intermediate stage, the cognitive skills acquired during the student years as its output, with the number of years of schooling and the quality of the education received by the student years as its input; and at the most basic stage, the quantity of education and the quality of education as its output. In the most basic stage, the quality and quantity of education are outputs, while per capita expenditure, enrolment, teacher-student ratios, etc., are inputs.**

(ii) Classification and characteristics of the various human capital measures

Analyze the various human capital measurement indicators under the framework. Before forming the framework, there is a need to first summarize the indicators that emerged.

1. Outputs - inputs, direct - indirect, stocks - flows, quality - quantity

We have summarized these indicators mentioned by direct-indirect, stock-flow, input-output and quality-quantity based on the keywords described earlier.

Table 8-1 Individual Human Capital Measures under the Direct-Indirect and Stock-Flow Classifications

	Stock indicators	Flow indicators
straightforward	Adult cognitive skills (mean)	Student cognitive skills (mean)
overhead	QAYS (all populations) Human capital separated from wage information Average years of schooling literacy	QAYS (a certain birth queue) Quality of education separated from wage information Teacher-student ratio, per pupil funding, enrollment, etc.

Table 8-2 Measures of Human Capital by Input-Output, Quality-Quantity Divisions

arrangement of ideas	Inputs OR outputs	Quantitative indicators	Quality Indicators

the top of a building	outputs	Adult cognitive skills (mean)	
	throw oneself into	Cognitive skills acquired by adults as students, QAYS (a certain birth cohort)	
middle-ranking	outputs		
	throw oneself into	Educational quality, (grade or grade-consistent) student cognitive skills separated from wage information (mean)	Average years of schooling
demersal	outputs		
	throw oneself into	Teacher-student ratio, per pupil funding	enrolment

2. Labor market and education system

While all of the above indicators can be used to measure a country's human capital, in terms of their bias, some of the indicators are more inclined to measure indicators of a country's education system. In the terminology used in the text, we have tried to make a distinction between human capital and education quality, and in the table below, we summarize the preferences of these indicators. The important and necessary reason for this distinction stems from the focus on information in the time dimension, where the impact of the education system on overall human capital is gradual, slow and lagging, and therefore, if there is to be a correspondence, the education system indicators pre-date the human capital measurements, which is ignored by some of the treatments in the existing studies.

We argue that average years of schooling is both a human capital measure and an education system indicator, because here education is used purely to measure human capital; student cognitive skills are an education system indicator because they are measured mostly for students in primary and secondary education; adjusted average years of schooling, based on whether average years of schooling are for the population as a whole or for the birth cohort that has just entered the labor market population, which are biased towards human capital and the education system, respectively, but it is important to know that the quality of education used for the adjustment is theoretically obtained from the data on the cognitive skills of the students, as the theory of the method should be biased towards the education system; the cognitive skills of the adults, since they are the cognitive skills of the population of the labor market, which, as advocated in this paper, are proxies for the optimality of the human capital; and the information on the wages of the population in the labor market, which is used to isolate the quality of education and therefore its also biased towards the education system.

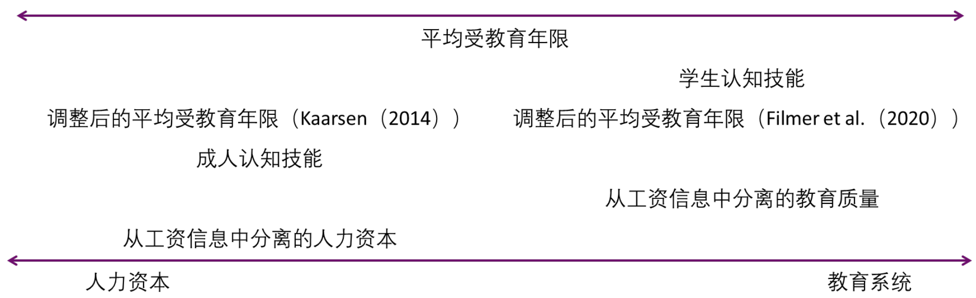


Figure 8-1 Schematic of the tendencies of each measure under the human capital-education system division

(iii) Individual human capital measures based on a harmonized framework

1. Principles for evaluating the strengths and weaknesses of human capital measurement indicators

The previous history of the evolution of human capital indicators would have reflected tendencies in the evaluation of human capital indicators, which are summarized in this paper in four evaluation principles: "direct indicators are preferred to indirect indicators", "stock indicators are preferred to flow indicators", "Output indicators are preferred to input indicators", and "Quality indicators are preferred to quantity indicators".

First, if there is the so-called optimal measurement of human capital, it must be directed to human knowledge and skills, and any measurement indicator that does not involve knowledge and skills is only an indirect measurement of human capital, which means that "direct indicators are better than indirect indicators". From this perspective, if we want to measure the human capital of a country's education, it is sufficient to measure the cognitive skills of adults in that country; if we want to investigate the quality of a country's education system, it is sufficient to measure the change in the value-added of the cognitive skills of the country's schoolchildren before and after they receive formal education.

Second, human capital can contribute to economic growth, and although not explicitly specified, human capital in this context should be the human capital of all people, and should therefore, in theory, be measured directly for all adults in the labor market (i.e., stock measurements) rather than for students (i.e., flow measurements). The impact of a country's educational development on the human capital of its population is lagged. Established studies have also confirmed that basic education students' cognitive skills and adults' cognitive skills usually change significantly in distribution in most countries in the world, so the practice of using current students' cognitive skills as a proxy for adults' cognitive skills over the same period is questionable (Bin Huang et al., 2024). Thus "stock indicators are preferred to flow indicators".

Third, in the **dynamic process of** transforming **educational inputs and outputs at different levels**, although various input and output indicators have been used to measure a country's educational human capital. However, we should be aware that the measurement of outputs should rightly take precedence over the measurement of inputs, as it is not appropriate to consider all inputs as effective productive investments (Hanushek, 2003), and there is uncertainty as to how much outputs the inputs can produce (Schoellman, 2012; Hanushek, 2003). In this case, it is more correct to measure outputs directly, so that "output indicators are preferable to input indicators".

Thirdly, with the widespread dissemination and acceptance of human capital theory, globally, almost every country develops education and human capital. The development of education should focus on both the quality of education and the quantity of education, both of which are aimed at improving the final cognitive skills output. If a country provides an extremely low quality of education to its population, even if the average number of years of education of its nationals reaches more than 16, the level of cognitive skills of its nationals will not be too high, and such an investment in education is inefficient and will not generate economic value for external socio-economic development. Similarly, if a country provides a limited amount of education to its people, even if the quality of education in that country is high, the level of cognitive skills of its nationals will not be too high, because the impact of one year of education cannot be too large, and the economic value of such an investment in education to the external economy and society will ultimately be limited. Therefore, for long-term national development, quantitative expansion of education and quality improvement are equally important. However, as the general idea of investing in human capital spreads globally, much of the message has changed and distorted, losing much of its essential content and power (Hanushek & Woessmann, 2015). Instead of really focusing on the essence of human capital: knowledge, skills, etc., policymakers and scholars have focused on proxies related to educational attainment, such as average years of schooling, school enrollment, and other quantitative dimensions. The reality of using quantitative indicators such as average years of schooling and enrollment rates as the focus of government policy tells us that some countries (e.g., Latin American countries) have expanded educational opportunities and increased average years of schooling with their own help and that of other countries, but their economic growth has remained slow and has not significantly caught up with the traces of developed countries (Hanushek & Woessmann, 2008). These indicators mask differences in the quality of education across countries; and "Schooling is not learning" (Pritchett, 2013; World Bank, 2018; Kaffenberger & Pritchett, 2017; Filmer et al., 2020; Angrist et al., 2021), it is only by transforming schooling into effective learning and the enhancement of real skills that educational development can be productive and thus have a sustained endogenous impact on a country's long-term economic growth (Huang, B. & Yun, R., 2023). Yun Ruxian, 2023). Therefore, in this context, "quality indicators are better than quantity indicators".

2. Analysis of human capital indicators

From the perspective of direct and indirect measurement, indicators of (student/adult) cognitive skills are superior to indicators such as QAYS or LAYS, years of schooling, etc.; from the perspective of stock and flow, indicators of adult cognitive skills and years of schooling are superior to indicators of student cognitive skills and enrollment; from the perspective of outputs and inputs, indicators such as indicators of adult cognitive skills are superior to indicators such as teacher-student ratios and per-pupil expenditures; and from the perspective of quality and quantity From the perspective of output and input, indicators such as the indicator of adult cognitive skills are superior to indicators such as teacher-student ratio and per pupil funding; from the perspective of quality and quantity, indicators such as the indicator of student cognitive skills are superior to indicators such as years of education and enrollment. In the light of the above comparisons, adult cognitive skills should be the best indicator for measuring the human capital of national education, and it has an absolute advantage over other indicators.

Second, we believe that if the goal is to measure human capital, priority should be given to labor market indicators, priority should be given to the use of direct measurements if they are available, priority should be given to stock measurements if there are no direct measurements, and priority should be given to output indicators if there are no stock measurements, and priority should be given to quality measurements among the input indicators.

(1) Revisiting the Relationship between Adult Cognitive Skills, Years of Education, and Students' Cognitive Skills

There is a need to restate what is known about the relationship between adult cognitive skills, average years of schooling, and student cognitive skills. These three are the focus of existing research, and increased awareness of them would help to understand existing research.

Based on the analytical framework, the following equations are available:

$$\text{成人认知技能} = f(\text{对应学生时代认知技能})$$
$$= g(\text{教育质量} * \text{教育数量}(\text{受教育年限}))$$

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

Adult cognitive skills are a function of the cognitive skills of adults corresponding to the student years. It is important to note that cognitive skills corresponding to the student era are vastly different from student cognitive skills; first, they do not coincide in time; the student era corresponding to today's adults is located many years ago, whereas today's student cognitive skills are of today; second, conceptually, the cognitive skills corresponding to the student era are the end result of his or her education, which is the ideal, and in fact is not measured If it is measured, it should be measured at the time of its final departure from school, which makes the matter difficult to implement due to the fact that different people leave school at different times; and the actual measurement of the cognitive skills of the students, for measurement reasons, refers more to the current students of a certain age or grade, who are still in education, and whose cognitive skills are not the final result.

Because of the above distinction, it follows that the following equation does not hold, strictly speaking:

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

Second, in the case of student cognitive skills, they can be used as the quality of education as long as the quantity of their education is consistent, i.e., they are assessed on students of the same age or grade, otherwise, student cognitive skills do not equal the quality of education.

IX. Summary, evaluation and outlook

Reliable and accurate measurement of human capital across countries is fundamental to macro-level education policy research. Along the "history of evolution" of human capital measurement, this paper systematically introduces human capital indicators such as years of schooling, students' cognitive skills, QAYS, adults' cognitive skills, and quality of education and human capital separated from labor market wage information, including their 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 types of indicators, analyze their respective problems, condense the basic principles for judging the advantages and disadvantages of various types of measurement indicators, and point out the direction of development for future research.

(i) Summary: patterns of development of human capital measurement indicators

Looking at the process of the development of human capital measurement indicators, the following patterns can be clearly found, first, the priority is to measure education acquisition or human capital by output indicators rather than input indicators, with average years of schooling and cognitive skills being the direct output-side indicators. Measuring the education system or human capital in terms of outputs rather than inputs is an outcome-oriented result, a response to the question, "Will the inputs lead to outputs? How much can they produce?" It is the result of thinking about "Can inputs produce outputs? How much output can they produce?

Second, there has been a shift from a focus on quantitative indicators to an increasing emphasis on qualitative indicators, which is well reflected in the shift in human capital measures from average years of schooling to cognitive skills. This is influenced both by the fact that existing policies aimed at quantitative indicators, such as average years of schooling, have not led to the expected economic growth, as well as by differences in the quality of education between education systems and the perception that schooling is not the same as learning.

Third, in macro-level studies, researchers have moved from considering only quantity to considering both quantity and quality of education. It should be noted, however, that for individuals, this means both a sufficiently large quantity of education and a sufficiently good quality of the education received; for countries, this means that "a large educated population with insufficient quality of education" and "excellent quality of education with a limited educated population" both limit the development of a country's productive capacity. For countries, this means "large educated population but insufficient quality of education" and "excellent quality of education but limited educated population", both of which limit the development of a country's productivity.

Fourth, while too much attention has been paid to flow indicators (student cognitive skills) in cognitive skills data in recent years, there is a tendency to return to stock indicators (adult cognitive skills). Human capital stock generally refers to the human capital of the entire labor force, which has a higher causal association with related indicators, such as the economy, relative to human capital flows (De La Fuente & Doménech, 2024).

(i) Evaluation: limitations of human capital measurement indicators and data

In the analysis in Chapter 8, it is assumed that each human capital indicator is perfect and its measurement is not problematic. However, it is also known from Chapters 3 through 7 that, due to data construction, the

1. Years of education

In order to obtain comparable average years of schooling across countries, harmonization and estimation are required in the construction of the data, for example, by harmonizing the different stages of education across countries into a few broad stages. Since the statistics published by each country are based on their own international normative standards, these operations may lead to discrepancies between the results of the construction and the statistics published by each country, and it is therefore not advisable to make direct comparisons between the constructed data and those published by each country. In addition, the biggest drawback, as it has been criticized, is that it does not reflect differences in the quality of education received.

2. Cognitive skills

The use of cognitive skills to measure the quality of education or human capital has become an important current research and development direction, however, it is undeniable that there are a number of limitations as far as the current data is concerned.

In terms of what is measured, there are currently no test items that cover and measure all the competencies and skills that determine a country's capacity for innovation and workforce productivity, including non-cognitive skills, skills acquired at university and in the workplace, and the highly specialized and complex knowledge and skills of scientists and high-level technologists (De La Fuente & Doménech, 2024). In terms of the target population to be measured, existing international cognitive skills testing programs measure more information about students' cognitive skills, and therefore can only be used to measure the quality of primary or secondary education; there are no data on cognitive skills testing programs that measure quality at the tertiary level; for countries with lower enrollment rates, these student cognitive skills testing program scores may represent only the birth cohort that participates in the testing program For countries with low enrollment rates, these student cognitive skills test program scores may only represent the knowledge and skills of a subset of the birth cohort participating in the test program, rather than the entire birth cohort; furthermore, students have not yet entered the labor market, and thus there are insufficient labor force substitutes to serve as a measure of the quality of education or human capital for the entire country, and there is a serious theoretical problem of endogeneity and reciprocal causation (Huang, B. et al. 2024; De La Fuente & Domé nech, 2024). nech, 2024)[58] . In terms of measuring countries and time, we are also still far from having enough information to measure human capital (and change) over time in most countries around the globe. Even for student testing programs, many countries have participated in only the most recent international testing programs, and thus only a few countries have data on long time series. In addition, the quality of the available cognitive skills data leaves much to be desired, and while studies have alluded to the risks that may exist when the assumption of identical content is not strictly met, there are no studies that have yet considered how to improve the quality of conversion data when the assumption of

[58] In this case, studies using the average number of years of schooling of the labor force population may be relatively less affected by the reverse causation problem (De La Fuente & Doménech, 2024). Of course, the most straightforward approach is still to use adult cognitive skills directly (Bin Huang et al., 2024).

identical content is not strictly met by addressing test content discrepancies in some way and obtaining conversion data that are more consistent in their connotations.

In the case of students' cognitive skills as educationally acquired for this use, scholars usually consider only comparable cognitive skill scores, ignoring student concordance (e.g., age).

3. Rehabilitation of average years of schooling

The transformation of the average years of schooling to include both quantitative and qualitative dimensions in the same indicator does not discard the commonly used average years of schooling on the one hand, but also adds cognitive skills as an indicator of the qualitative dimension on the other.

However, this indicator contains large construction errors. The implementation of the study relies on assumptions about the cognitive skill production function (which is undeniably better understood), and a lack of understanding of cognitive skill production will limit the accuracy of the final results obtained by this method; overcoming a lack of understanding of the cognitive skill production function will require measuring cognitive skills at each age group and observing generalized patterns of cognitive skill growth. Overcoming the lack of understanding of the cognitive skill production function will require measuring cognitive skills in each age group and observing generalized patterns of cognitive skill growth, which should take into account both educational influences and physical and psychological maturation, for which more effort is needed on the part of the researcher. Moreover, when using adult cognitive skills to estimate the quality of education, the presence of depreciation in cognitive skills and the influence of factors such as on-the-job training, dry schooling, and so on (De La Fuente & Doménech, 2024) make it more difficult to obtain a clean picture of the quality of the education system, with greater errors in the estimation. In addition, it is easy to overlook that cognitive skills are similarly subject to measurement error in their measurement, which accumulates to the average number of years of schooling after retrofitting.

Second, the usefulness of this indicator is limited. When the quality of education is consistent across educational stages, the cognitive skills score of students in a given grade is a better indicator if one wants to measure the quality of education; when the quality of education is inconsistent across educational stages, the value-added change in students' cognitive skills scores is a better indicator if one wants to measure the quality of education; when the quantity of education is measured, the average number of years of schooling is a better indicator; and when measuring the final educational output, the final adult cognitive skills score is the best indicator. In this framework, there is no doubt that there is no need to revamp the average years of schooling, and in order to better serve policy goals, education statistics and surveys should be strengthened to accurately capture data on the years of schooling of the population, as well as data on the cognitive skills of the population in each age group.

4. Wage information on the market

While isolating education quality or human capital from wage information in the labor market is feasible when data on immigrants are available, the specificity of this approach puts it at a disadvantage when comparing it with data on average years of schooling and cognitive skills, resulting in its limited use in measuring education quality or human capital.

First, in terms of considerations, the use of wage information from the labor market makes it necessary for the study to also take into account a number of factors that occur after the student enters the labor market, which makes it theoretically more difficult to implement this type of study than previous ones.

Second, the use of data in practice is more demanding. On the one hand, this study requires the use of migration data, which is inherently more difficult to obtain; and its results are susceptible to the influence of different sources of data, although there is a strong correlation between the results obtained with migration data from different countries, it is not possible to deny the instability of the results of . On the other hand, the quality of the education system it measures is the weighted value of individuals entering the labor market, and since such studies assume that the quality of the education system is constant, thereby neglecting to take time into account, ideally they should include a sufficiently representative sample of individuals from all periods. If the sample of individuals in some of the countries observed is early entrants to the labor market, their market returns may measure the quality of the education system in that country at an earlier time, and may be practically incomparable with those in other countries.

Finally, in terms of time-variation, the indicators obtained by such studies do not measure changes in education quality and human capital over time. Such studies are all based on data on the earnings of immigrants with different graduation times and similar working times to obtain an estimate of the quality of education or human capital in a country that does not vary over time. However, the quality of education in a country varies across time, making the individual returns to the same amount of education, as well as the human capital gained from education across time, different, while this approach makes it difficult to compare differences in the quality of the education system or human capital over time.

These limitations make it difficult to use the quality of education or human capital from this type of research in policy practice.

(ii) Looking forward: recommendations for the use of human capital measurement indicators

First, a distribution will always contain more information than a single mean. This is true for both educational attainment and cognitive skills. In educational attainment, the distribution of educational attainment (the share of the population that has reached a certain level of education) is more informative than a single mean number of years of schooling; in cognitive skills, the distribution of cognitive skills (mean, quantile, standard deviation, skewness) is more informative than a single mean of cognitive skills. The emphasis on distributions, although driven by Hanushek & Woessmann (2012a), Bin Huang and Ruxian Yun (2023), and Bin Huang et al. (2024), has yet to be explored in depth. The use of distributional information is both a full utilization of existing information and an innovative development of human capital theory in empirical evidence, implying an extension of the study from the simple overall level of human capital to richer connotations such as differential characteristics of human capital and structural characteristics. In addition, the applicability and accuracy of various methods applied to the transformation of other distributional characteristics in the construction of comparable databases of students' cognitive skills need to be further analyzed and discussed.

Second, changes in the quality of education over time should not be ignored. In the context of rapid educational expansion, especially in higher education, it is particularly important to study changes in human capital over time. The quality of education remains constant is the basic assumption of most studies (De La Fuente & Doménech, 2024), however, during periods of rapid educational expansion, both the quantity and quality of education change dramatically, and not considering the time dimension in the measurement of human capital at this time will affect the reliability of the results of the study; simple empirical analyses have also shown that, in many countries, education quality has not remained stable over time (Hanushek & zhang, 2009). It is now possible to discuss changes in educational quality during periods of rapid expansion by considering information on adult cognitive skills and educational attainment across different birth cohorts.

Third, while adult cognitive skills are an optimal measure of a country's human capital, the use of cognitive skills as a measure of education quality or human capital needs to be further understood. Neither student cognitive skills nor adult cognitive skills are the product of schooling alone, and no study has yet been able to provide an accurate answer to the question of how much the influence of out-of-school education contributes to the impact of cognitive skills. In addition, while research often uses student cognitive skills as a measure of the quality of education in each country, the production of student cognitive skills should also be a joint output of both quantitative and qualitative inputs to education, and therefore the cognitive skills of students at the same grade level (at which point the same number of years of education have been received) should be used, rather than a mixture of data from multiple grade levels. While recognizing this point does not detract too much from previous research, the emphasis on this is not common.

bibliography

[1] Huang, B., Yun, R.. (2023). What makes educational development a strong country - an empirical analysis based on internationally comparable data on cognitive skills from 1960-2020. Educational Research, 44(10), 125-136.
[2] Huang B, Yun Ruxian, Wu Kailin. (2024). The impact of cognitive skill distribution on national economic growth: new evidence from educationally strong countries. Journal of East China Normal University (Education Science Edition), 42(9), 13-32.
[3] Altinok, N., Angrist, N., & Patrinos, H. A. (2018). Global Data Set on Education Quality (1965-2015). *Policy Research working paper*, Washington, D.C.: World Bank Group.
[4] Altinok, N., & Diebolt, C. (2024). Cliometrics of Learning-Adjusted Years of Schooling: Evidence from a New Dataset. *cliometrica,* 18(3), 691-764.
[5] Altinok, N., Diebolt, C., & Demeulemeester, J. L. (2014). A New International Database on Education Quality: 1965-2010. *applied Economics*, 46(11), 1212-1247.
[6] Altinok, N., & Murseli, H. (2007). International Database on Human Capital Quality. *economics Letters*, 96(2), 237-244.
[7] Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring Human Capital Using Global Learning Data. *nature*, 592(7854), 403-408.
[8] Angrist, N., Evans, D., Filmer, D. P., Glennerster, R., Rogers, F. H., & Sabarwal, S. (2020). How to Improve Education Outcomes Most Efficiently? A Comparison of 150 Interventions Using the New Learning-Adjusted Years of Schooling Metric. CDG CDG *Working Paper*.
[9] Balaj, M., Henson, C. A., Aronsson, A., et al. (2024). Effects of Education on Adult Mortality: a Global Systematic Review and Meta-Analysis. The Lancet Public Health, 9(3), e155-e165.
[10] Barro, R. J., & Lee, J. (1993). International Comparisons of Educational Attainment, *Journal of Monetary Economics*, 32(3), 363-394.
[11] Barro, R. J., & Lee, J. (2001). International Data on Educational Attainment: Updates and Implications. *Oxford Economic Papers*, 53(3), 541-563.
[12] Barro, R. J., & Lee, J. (2013). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 104(September), 184-198.
[13] Barro, R. J., & Lee, J. (2015). Education Matters: global schooling Gains from the 19th to the 21st Century. Oxford, UK: Oxford University Press.
[14] Bauer, R., Potančoková, M., Goujon, A., & K.C., S. (2012). Populations for 171 Countries by Age, Sex, and Level of Education around 2010: Harmonized Estimates of the Baseline Data for the Wittgenstein Centre IIASA Interim Report IR-12-016.
[15] Braun, H. I., & Holland, P. W. (1982). Observed-score test equating: a mathematical analysis of some ETS equating procedures. In P. W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 9-49). New York: Academic.
[16] Cohen, D., & Soto, M. (2007). Growth and Human Capital: Good Data, Good Results. *Journal of Economic Growth*, 12(1), 51-76.
[17] Das, J. & Zajonc, T. (2010). India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement. *Journal of*

Development Economics, 92(2), 175-187.
[18] De La Fuente, Á., & Doménech, R. (2000). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make? *Working Paper*.
[19] De La Fuente, Á., & Doménech, R. (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make? *Journal of the European Economic Association*, 4(1), 1-36.
[20] De La Fuente, Á., & Doménech, R. (2015). Educational Attainment in the OECD, 1960-2010. Updated Series and a Comparison with Other Sources. *economics of Education Review*, 48(October), 56-74.
[21] De La Fuente, Á., & Doménech, R. (2024). Cross-Country Data on Skills and the Quality of Schooling: a Selective Survey. *Journal of Economic Surveys*, 38(1), 3-26.
[22] Égert, B., De La Maisonneuve, C. & Turner, D. (2024). A New Macroeconomic Measure of Human Capital Exploiting Pisa and Piaac: Linking Education Policies to Productivity. *Education Economics*, 0(0), 1-17.
[23] Filmer, D., Rogers, H., Angrist, N., & Sabarwal, S. (2020). Learning-Adjusted Years of Schooling (LAYS): Defining a New Macro Measure of Education. *Economics of Education Review*, 77(August), 101971.
[24] Gethin. (2023). Distributional Growth Accounting: Education and the Reduction of Global Poverty, 1980-2022. Job Market Paper.
[25] Glawe, L., & Wagner, H. (2022). Is Schooling the Same as Learning? - The Impact of the Learning-Adjusted Years of Schooling on Growth in a Dynamic Panel Data Framework. *World Development, 151(2022)*, 105773.
[26] Goujon, A., K.C., S., Speringer, M., et al. (2016). A Harmonized Dataset on Global Educational Attainment Between 1970 and 2060 - an Analytical Window into Recent Trends and Future Prospects in Human Capital Development. *Journal of Demographic Economics*, 82(3), 315-363.
[27] Gust, S., Hanushek, E. A., & Woessmann, L. (2024). Global Universal Basic Skills: Current Deficits and Implications for World Development. *Journal of Development Economics*, 166(January), 103205.
[28] Hastedt, D. & Desa, D., (2015) "Linking Errors Between Two Populations and Tests: a Case Study in International Surveys in Education ". *Practical Assessment, Research, and Evaluation* 20(1), 14.
[29] Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal*, 113(485), 64-98.
[30] Hanushek, E. A., & Kimko, D. D. (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review,* 90(5), 1184-1208.
[31] Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3), 607-668.
[32] Hanushek, E. A., & Woessmann, L. (2012a). Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. *journal of Economic Growth*, 17(4), 267-321.
[33] Hanushek, E. A., & Woessmann, L. (2012b). Schooling, Educational Achievement, and the Latin American Growth Puzzle. *Journal of Development Economics*, 99(2), 497-512.

[34] Hanushek, E. A., & Woessmann, L. (2015). Universal Basic Skills: what countries Stand to Gain. Organisation for Economic Co-operation and Development, Paris.
[35] Hanushek, E. A., & Zhang, L. (2009). Quality-Consistent Estimates of International Schooling and Skill Gradients. *Journal of Human Capital*, 3(2), 107-143.
[36] Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics*, 25(2), 133-183.
[37] Kaarsen, N. (2014). Cross-Country Differences in the Quality of Schooling. *Journal of Development Economics*, 107(March), 215-224.
[38] Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics,* 9 (1) , 25-44.
[39] Kolen, M. J., & Brennan, R. L. (2014). Test Equating, Scaling, and Linking: Methods and Practices. new York: Springer.
[40] Krueger, A. B., & Lindahl, M. (2001). Education for Growth: Why and for Whom? *Journal of Economic Literature*, 39 (4), 1101-1136.
[41] Lee, J., & Barro, R. J. (2001). Schooling Quality in a Cross-Section of Countries. *Economica*, 68(272), 465-488.
[42] Lim, S. S., Updike, R. L., Kaldjian, A. S., et al. (2018). Measuring Human Capital: a Systematic Analysis of 195 Countries and Territories, 1990-2016. *the Lancet*, 392(10154), 1217-1234.
[43] Lucas, R. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22(1), 3-42.
[44] Lutz, W., Goujon, A., K. C., S., & Sanderson, W. C. (2007). Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000. *Vienna Yearbook of Population Research*,. 193-235.
[45] Martellini, L., & Schoellman, T. (2012). Human Capital and Development Accounting: New Evidence from Wage Gains at Migration. *the Quarterly Journal of Economics*, 133(2), 665-700.
[46] Patel, D., & Sandefur, J. (2020). A Rosetta Stone for Human Capital. *cgd Working Paper*.
[47] Patrinos, H. A., & Angrist, N. (2018). Global Dataset on Education Quality: a Review and Update (2000-2017). *Policy Research working paper*, Washington, D.C.: World Bank Group.
[48] Pritchett, L. (2013). The Rebirth of Education: Schooling ain't Learning. Washington, D.C.: CGD Books.
[49] Reardon, S. F., Kalogrides, D., & Ho, A. D. (2021). Validation Methods for Aggregate-Level Test Scale Linking: a Case Study Mapping School District Test Score Distributions to a Common Scale. *Journal of Educational and Behavioral Statistics*, 46(2), 138-167.
[50] Reiter, C., Özdemir, C., Yildiz, D., Goujon, A., Guimaraes, R., & Lutz, W. (2020). The Demography of Skills-Adjusted Human Capital. *working Paper*.
[51] Romer, P. (1986). Increasing Returns and Long-run Growth. *Journal of Political Economy*, 96(5), 1002-1037.
[52] Sandefur, J. (2018). Internationally Comparable Mathematics Dcores for Fourteen

African Countries. *economics of education review*, 62(February), 267-286.
[53] Schoellman, T. (2012). Education Quality and Development Accounting. *The Review of Economic Studies*, 79(1), 388-417.
[54] Schultz, T. M. (1961). Investment in Human Capital. *American Economic Review*, 51(1), 1-17.
[55] Singh, A. (2014). Emergence and Evolution of Learning Gaps across Countries: Linked Panel Evidence from Ethiopia, India, Peru and Vietnam. *csae working paper*.
[56] Speringer, M., Goujon, A., K.C., S., Potančoková, M., Reiter, C., Jurasszovich, S., & Eder, J. (2019). Global Reconstruction of Educational Attainment, 1950 to 2015: methodology and assessment. *vid Working Paper*