ARTICLE

Check for updates

Convergence of primary education development in urban and rural China: empirical analysis of historical trends and future projections

Yan Zhang[1], Wei Yang[2], Bo Li[3], Yifan Yang[1], Liu Chen[1] & Lu Feng [1✉]

Integrating primary school education in urban and rural areas is essential for reducing disparities in resources and promoting educational equity. This study uses four analytical methods—entropy variation coefficient-comprehensive index, time difference correlation analysis, multivariate linear regression prediction, and Beta convergence—to analyze the historical evolution, future trends, and convergence in primary education development levels in China. Using national time series data from 1995 to 2022, we constructed an evaluation index for primary education and found that the urban-rural gap in primary education has slightly narrowed. Although urban growth currently surpasses rural growth, projections for the next 13 years suggest a decreasing gap in annual growth rates between the two. The analysis shows significant absolute Beta convergence nationally and within urban areas from 1995 to 2035, while rural areas exhibit divergence. After controlling for factors such as per-student expenditure, school building floor space per student, student-teacher ratio, teacher education level, teacher title structure, number of primary schools per student, enrollment rates, and dependency ratio, we find no significant conditional convergence at the national level or between urban and rural areas from 1995 to 2035. These results highlight that current levels of urban-rural investment are inadequate for balancing primary education development. To address this, increased and equitable educational investment is necessary for balanced development across urban and rural areas, within rural areas, and among urban regions.

[1] College of Economics and Management, Yunnan Agricultural University, 650201 Kunming, Yunnan, China. [2] School of Agriculture and Environment, Massey University, Manawatu campus, Palmerston North 4410, New Zealand. [3] International Agricultural Research Institute, Yunnan Academy of Agricultural Sciences, 650205 Kunming, Yunnan, China. ✉email: fenglu1128@126.com

Introduction

According to statistics from the World Bank for 2022, the global population reached 7.837 billion by the end of 2021, with 43.4% residing in rural areas. High-income countries have 18.5% of their population living in rural areas, while the majority of rural populations are found in middle-income, low-middle-income, and low-income countries, particularly in low-income countries where rural residents constitute 65.8% of the population. Countries such as India, Cambodia, Myanmar, Laos, Pakistan, Bangladesh, Sri Lanka, and Vietnam have a rural population exceeding 60%. Moreover, the global poverty rate in rural areas has reached 17% (compared to 7% in urban areas), with rural communities accounting for 70% of the world's extreme poverty, according to the Rural Culture Committee of China Cultural Management Association in 2024. The persistent disparity between urban and rural development is largely attributed to the long-standing "dual economy" in underdeveloped regions and inadequate public services, including education, infrastructure, healthcare, and sanitation in most rural areas.

Education, as a public or quasi-public good, serves as a fundamental driver of human progress and social development. Country-specific experiences underscore the importance of balanced educational development in enhancing national comprehensive strength and fostering sustainable social progress (Yuan et al. 2020). Enhancing national quality and ensuring social equity through balanced primary education has become a universal policy focus for governments worldwide (Wang, 2018). The United Nations, in its 2030 Education Action Framework, emphasizes "developing fair and high-quality education" as a core educational goal, considering the significant urban–rural development imbalance globally, particularly in underdeveloped regions (UNESDOC, 2015). The European Commission has similarly prioritized the efficiency and equity of education systems (EU Commission, 2006). In India, the role of high-quality primary education is recognized as vital for unlocking economic potential, particularly in rural areas. Effective primary education for rural communities lays a stable foundation for higher education, skill acquisition, and employment opportunities. To this end, India advocates for corporate investment in social welfare initiatives aimed at improving rural education, reducing poverty, and supporting sustained economic growth (Ojha and Chakravarty, 2012; Shukla, 2020).

Renowned economist Theodore W. Schultz, in his book "Transforming Traditional Agriculture", emphasized that "neglecting primary and secondary education is very short-sighted."(Schultz, 1987). Similarly, Nobel laureate Douglas North believed that throughout history, institutional incentives for investing in productive knowledge have been lacking, often misdirecting resources toward higher education instead of primary education (North D C Institutions, 1990). According to the human capital theory, new structural economic theory, and the comparative advantage strategy of developing countries, primary education should be prioritized. This approach suggests that governments should assume greater responsibility for investing in primary and compulsory education to ensure equity and fairness. By adopting such a strategy, developing countries can alleviate income inequality, support economic development, and accelerate their progress toward parity with developed nations. Conversely, failure to prioritize primary education may hinder these countries' ability to transition into the middle-income bracket and risk falling into the middle-income trap (Krueger and Lindahl, 2001; Lin, 2008; Cao and Li, 2018).

China has long prioritized primary education, as evidenced by its efforts to make compulsory education widely accessible, achieve balanced development, and pursue high-quality and balanced compulsory education. These actions reflect the country's strategic reforms and adjustments. By the end of 2021, 2895 county-level administrative units in China have achieved the national recognition of basic educational balance (Meng and Yao, 2022). Despite efforts to enhance the urban-rural education balance through measures such as increased funding for rural compulsory education, improved rural school infrastructure, the "two exemptions and one subsidy" policy, and better compensation and training for rural teachers, there is still an imbalance between urban and rural areas. Approximately 500 million Chinese people still live in rural areas, and urbanization lags behind industrialized nations by 72.6% (Cai et al. 2019; Tang et al. 2022). Contributing factors include the entrenched dual economic system, rural-to-urban migration, unequal economic and social development, and imbalanced educational resource allocation. Consequently, rural schools, especially primary schools, often face challenges such as being under-resourced yet underpopulated, while urban primary schools are well-resourced but overcrowded (Xie, 2014). These issues hinder China's efforts toward rural revitalization and urban-rural integration.

Currently, in the research on the development gap between urban and rural education, one is to take a qualitative approach, such as urban and rural education integration (Zheng et al. 2023; Tang et al. 2023; Lei and Xie, 2023), education equity (Jia, 2023), high-quality and balanced rural compulsory education development (Lu et al. 2023), digital transformation of rural basic education (Hu et al. 2023), and rural primary education system evaluation (Singh and Parashar, 2024), etc. The second is to take quantitative exploration, such as the resource allocation in urban and rural primary education (Zheng and Yue, 2017; Feng and Zhang, 2018), migrants' primary education (Zhang et al., 2015; Zhang, 2017) and the role of compulsory education in poverty reduction (Li et al. 2019; Li, 2019; Zuo et al. 2023). Despite, various studies indicating that primary education in urban and rural China remains imbalanced, many suggest that it will become more balanced and eventually converge over time (Han and Li, 2013; Li and Luo, 2015; Yu and Wang, 2022; Xu and Liu, 2023). The emphasis in educational forecasting lies in estimating the extent of development (Xu and Liu, 2023), predicting academic performance (Field et al. 2019), school-age population changes (Zhou, 2024), and dropout rates (Nascimento et al. 2022). However, to our knowledge, no existing research predicts future trends or tests convergence in urban and rural primary education based on historical evolution.

The purpose of this study is to assess the development level of primary education in China and to create an evaluation index system for both urban and rural areas. Annual time series data were utilized on national primary education development from 1995 to 2022 (a complete data set available at the national level). The coefficient of variation entropy-comprehensive index method is first employed to identify the historical evolution characteristics of urban and rural primary education development in China. Secondly, time difference correlation analysis is used to screen early warning indicators and identify the leading factors influencing primary education development in urban and rural areas. Thirdly, a multiple linear regression model is applied to predict the future development of primary education from 2023 to 2035. Finally, Beta convergence analysis examines past and projected future convergence in primary education development between urban and rural areas. This study explores the disparity in primary education development between urban and rural regions, comparing their differences and evaluating the overall development level. This study provides technical methods and decision-making support for measuring and evaluating the development level of primary education in both urban and rural areas.

This study makes two key contributions to understanding the development of primary education in urban and rural China. Firstly, it's an initial attempt to apply the time difference correlation analysis method for forecasting the future development of primary education in urban and rural areas from 2023–2035, further exploring the convergence potential between urban and rural education. Secondly, this study introduces predictive warning models to the education field, offering a novel approach that may complement and expand existing research, thus enriching the current literature. Meanwhile, this paper highlights the issue of resource imbalance in rural primary education and predicts future development gaps in China, providing an early warning for educational challenges in similar regions.

The paper is structured as follows: Section "Literature review and theoretical framework" reviews the literature on the development of primary education and the construction of the index system. Section "Data and methodology", presents the data and research methodology. Including the coefficient of variation entropy-comprehensive index method, time difference correlation analysis method, multiple linear regression prediction model, and Beta convergence analysis method. Section "Empirical results", discusses the empirical results and findings. The paper concludes in section "Conclusion and discussion" and outlines limitations and future research directions in section "Limitations and future research".

Literature review and theoretical framework

Integration of urban–rural primary education. The integration of urban and rural primary education is essential for achieving equalization, fairness, coordinated and integrated development, and balanced linkage between urban and rural areas. While there is no universally agreed-upon definition of urban–rural primary education integration, interpretations may vary among different stakeholders' perspectives. Based on existing literature, this integration involves positioning urban and rural primary education within a unified system that dismantles the traditional binary opposition between the two. It emphasizes coordinated development, shared educational resources, the rational flow of educational elements, and the establishment of policies, institutional mechanisms, and resource allocation strategies to promote joint advancement (Chu, 2009; Zhang, 2012).

The integration does not imply making urban and rural primary education identical. Instead, it should be understood based on three aspects: consistency in educational goals, synergy in development speed, and stability and convergence of educational quality. Urban primary education typically advances at a faster rate with higher quality compared to rural areas. Bridging the gap requires accelerating rural primary education to catch up with urban progress, aiming for overall coordinated development. In terms of quality, it should integrate education management, resources, teacher management, and standardized evaluation across urban and rural areas. Additionally, there should be differentiation in teaching methods and content should be maintained to respect unique characteristics, alongside financial support directed at rural areas (Yang, 2019; Yan, 2020).

Numerous studies emphasizes that the integration should start with equalizing the supply of primary education in urban and rural areas. This can be achieved through enhanced financial mechanisms, equitable resource allocation, and targeted policy implementation. Moving away from a historically "urban preference" policy bias and adopting "rural preference" policies can help ensure a fairer distribution of resources, reducing the developmental gap (Shi, 2007; Yuan, 2011). Additionally, some scholars advocate for a symbiotic approach involving government, schools, families, and communities to create supportive environments for urban–rural education integration (Wu and Wang, 2022). However, existing studies have largely focused on optimizing supply-side factors, often overlooking the demand side. This study aims to address this gap by developing an indicator system that considers both the supply and demand aspects of primary education development in urban and rural areas.

Allocation of urban–rural primary education resources. Educational equity reflects social equity within the educational field, and resource allocation is a critical starting point for achieving educational equity (Wen and Gu, 2017). The allocation of educational resources is categorized under educational supply (Zhao et al. 2018). Research has extensively explored policies, influencing factors, and the effects of resource allocation in primary education across urban and rural areas. For example, Assaad and Krafft (2015) examined Egypt's free basic education policy, finding that issues of low quality and inefficiency often led families to incur significant costs for their children to succeed academically. Similarly, Molina-Muñoz et al. (2023) found that Spanish students' average mathematics scores fell below EU and OECD averages, attributing differences to socioeconomic status, gender, immigration status, language is spoken at home, and student–teacher ratios.

Evaluation index systems have been developed to measure the allocation of educational resources, considering factors such as teacher availability, funding, infrastructure, and school conditions. Researchers use different analytical methods, including PLS structural equation modeling and data envelopment analysis (DEA) have been employed to assess the efficiency and spatial distribution of resources. For example, Ramzi et al. (2016) applied DEA to assess the efficiency of primary and secondary education. They find that inefficiency in education is strongly related to poverty within governorates. Research by Chinese scholars highlights that primary schools in China are mainly concentrated in the east of the "Hu Huanyong Line", which has significant regional differences in educational resources. Resource allocation is generally higher in the eastern region compared to the central and western regions, where low comprehensive efficiency is often due to inadequate scale efficiency (Ma et al. 2020; Li and Lei (2021); Cai and Zhang, 2023). Notably, focusing solely on resource allocation without considering the demand side overlooks critical aspects of primary education development.

Evaluation indicators of urban–rural primary education development. Past studies have largely measured primary education development by focusing on educational conditions, teaching staff, and educational resources. For example, Cai et al. (2022) constructed an index system featuring four primary indicators—educational opportunities, educational conditions, educational quality, and educational equity—and 13 secondary indicators to assess the level of education modernization in ethnic regions. In contrast, Mo et al. (2022) replaced "educational conditions" with "educational investment" while maintaining the other three indicators. However, few studies have incorporated social and economic indicators such as per capita GDP, despite their significant impact on primary education development in urban and rural areas. This study posits that both the supply and demand sides are crucial for understanding education development. An increase in educational investment, teacher allocation, and economic indicators such as GDP per capita and disposable income per capita positively impact education quality and quantity. Therefore, a framework of indicator systems including education supply, education demand, and education quality has been constructed based on existing literature, the "National Basic

Equilibrium Evaluation Plan for Compulsory Education Development", and our thoughtful analysis of the study. The framework includes 16 sub-indicators organized into four subsystems: educational conditions, teacher strength, student development, and socio-economic conditions. In the following subsections, we will explain the reasons behind their inclusion in the model.

Educational conditions. The subsystem comprises six sub-indicators to measure primary education quality: per-student education expenditure, per capita output value of fixed assets, floor space of school buildings per student, books per student, computers per student, and the percentage of schools with a campus network. Per-student education expenditure reflects the level of financial investment in primary education and indicates governmental support levels. Investment in educational material resources is fundamental for educational development, and the richer the investment in educational material resources, the higher the level of educational development. Therefore, the sub-indicators encompass both hardware (e.g., floor space and fixed assets) and software resources (e.g., books and computers per student, campus network availability), reflecting overall educational resources access.

Teacher strength. This subsystem includes four sub-indicators: teacher–student ratio, the proportion of full-time teachers, average years of teacher education, and teacher title structure. Teachers are the essential human resources for education, and numerous studies have shown that the quality of teachers is a key factor in determining the quality of education. Small-class teaching fosters better academic outcomes, and a higher teacher–student ratio, more full-time teachers, and greater teacher professional titles contribute to stronger educational performance.

Student development. This subsystem encompasses three sub-indicators: the number of primary schools per student, the number of classes per student, and the enrollment rate. The aim of education is student development. Liu and Fang (2017) noted that areas with concentrated schools and student populations benefit from resource consolidation. The smaller class sizes enhance teacher-student interaction, promoting overall student development. The enrollment rate is another indicator of demand-side and the quality of school teaching and should be included in this subsystem.

Socioeconomic conditions. This subsystem includes three sub-indicators: per capita GDP, per capita disposable income, and the dependency ratio. The income disparities between urban and rural areas affect the investment in education. Based on existing research by Sun et al. (2023), an increasing rural household income significantly promotes their family investment in education, narrowing the urban–rural gap. Conversely, a higher dependency ratio signifies a greater economic burden on families, potentially reducing education investment.

Data and methodology
Data sources. This study primarily utilized data sourced from the *China Education Statistics Yearbook, China Educational Finance Statistics Yearbook, China Statistical Yearbook,* and *China Population and Employment Statistics Yearbook* for the period spanning 1995–2023. The data represent a comprehensive national sample. To facilitate a comparative analysis between urban and rural areas, the national data set was divided into two distinct groups: the urban sample and the rural sample. In instances where data points were missing for specific years, linear interpolation was applied to fill these gaps, ensuring the continuity and integrity of the data set.

Research methodology
Entropy of variation coefficient-comprehensive index method. The method of entropy variation coefficient-comprehensive index calculates the weights of each index using a combination of the coefficient of variation method and the entropy value method. It then applies the composite index method to measure the level of development in primary education.

The entropy method is an objective assignment method that determines weight based on the information provided by the indicators. The smaller the information entropy of the indicators, the greater the degree of variability, and the greater the weight. However, this method does not consider the interactions between indicators, the difficulty of handling missing values, and sensitivity to extreme values, which may lead to the distortion of the weight. The coefficient of variation is a statistical measure that is obtained by dividing the standard deviation by the mean of a data set. It is useful because it eliminates the issue of scale, objectively assigns values to each data point, and avoids subjectivity that may arise in other methods such as AHP (Zhang and Jia, 2016). And this method can reflect the degree of data dispersion but also consider the average level of data, so to make up for the lack of entropy method. The method of coefficient of variation is insensitive to the distribution shape of data and can deal with the extreme value and missing value better, so it can be used as a supplement to the entropy method in some cases.

Based on the coefficient of variation method having a strong compensating effect on the entropy method, a combination of the two methods was used after weighing their respective advantages and disadvantages. The weights obtained from the coefficient of variation method and entropy value method were utilized for this purpose. Since the application of these two methods has been established, the calculation of the combined weights will be explained in detail here. The weights of each indicator determined using the two methods will be set as the combination weights as v_i and τ_i, respectively, and i as each indicator, with a total of 16 sub-indicators in this study. According to the principle of minimum information entropy, the weights determined by the coefficient of variation method and entropy value method are coupled to obtain the combined weight ω_i.

$$\min \eta = \sum_{i=1}^{m} \omega_i \left(\ln \omega_i - \ln v_i \right) + \sum_{i=1}^{m} \omega_i \left(\ln \omega_i - \ln \tau_i \right)$$

$$\sum_{i=1}^{m} \omega_i = 1$$

$$\omega_i > 0$$

η is the objective function of minimum information entropy and m is the number of indicators. The above function is obtained by optimizing it according to the Lagrange multiplier method:

$$\omega_i = \frac{\sqrt{v_i \tau_i}}{\sum_{i=1}^{m} \sqrt{v_i \tau_i}}$$

Drawing on Ding Jianjun's method of calculating the Comprehensive Development Index (Ding, 2014), the level of development of primary education in overall, urban and rural areas from 1995 to 2022 in this study was calculated. The calculation formula is as follows:

$$E = \sum_{i=1}^{m} I_i \omega_i$$

where E denotes the level of development of primary education, I_i

is the standardized value of the i indicator in the sub-indicator, and ω_i is the weight of the i indicator.

Time difference correlation analysis method. Time difference correlation analysis is a statistical method used to investigate the relationship between two sets of data over time. It looks at the correlation between the two sets of data, which is a measure of how closely related they are. The correlation coefficient is used to quantify this relationship, and it ranges $[-1, 1]$. A value of -1 indicates a perfectly negative correlation, while a value of 1 indicates a perfectly positive correlation (Liu and Liu, 2017). The principle of its calculation is as follows:

Let the time-difference correlation coefficient be λ, the benchmark indicator is the vector of indexes of the level of development of primary education from 1995 to 2022, i.e. $Y = \{Y_1, Y_2, \cdots, Y_n\}$, the selected indicators are the sub-indicators in Table 1, and the vector of sub-indicators is $X_i = \{X_{i1}, X_{i2}, \cdots, X_{in}\}$, of which $i = 1, 2, \cdots, m$, m is the number of sub-indicators, and there are a total of 16 sub-indicators in the present study, $m = 16$; n ($j = 1, 2, \cdots, n$) is the total number of years, and the present study is for a total of twenty years from 1995 to 2022, and $n = 28$.

$$\lambda = \frac{\sum_{j=j'}^{n_t} (X_{i,j+t} - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=j'}^{n_t} (X_{i,j+t} - \bar{X})^2 \sum_{j=j'}^{n_t} (Y_j - \bar{Y})^2}}, t = 0, \pm 1, \pm 2, ..., \pm T, j'$$

$$= \begin{cases} 1, t \geq 0 \\ 1 - t, t < 0 \end{cases}$$

where t is the overrun or lag period, $t > 0$ is the overrun period and $t < 0$ is the lag period, so t is called the time difference or delay number. T indicates the maximum number of delays and n_t is the amount of data after alignment. When selecting indicators, usually calculate the correlation coefficient of time difference for a number of different delays, and take the absolute value of the largest time difference correlation coefficient $\lambda_\lambda = \max_{-T \leq t \leq T} \lambda_t$ as the correlation coefficient of the time difference between the selected indicator and the benchmark indicator, and the corresponding j' is the over-advanced or lagged period, and the closer this coefficient is to 1 the more desirable it is in the test. In the specific calculation, if the absolute values of the maximum and second largest time difference correlation coefficients are close to each other, necessary adjustments can be made to the overrun and lag periods according to the actual situation of economic activities. Lagged and synchronous indicators do not have much influence on the benchmark indicators, and only the overshooting and stronger correlation indicators are the main influencing factors of the benchmark indicators, which are thus used as indicators of the subsequent multiple linear regression prediction models.

Multiple linear regression models. The study of the linear relationship between a dependent variable and multiple independent variables is known as a multiple linear regression model. With a group n ($n = 28$) of k independent variables, the relationship is obtained:

$$y_j = \alpha + \beta_1 x_{j1} + \beta_2 x_{j2} + \cdots \beta_k x_{jk} + \varepsilon_j, j = 1, 2, \cdots, n$$

where y_j is the level of development of primary education in nationwide (or urban and rural areas) in year j, $\alpha, \beta_1, \beta_2 \cdots, \beta_k$ is the parameter to be solved for, and ε_j is the random error.

The goodness of fit and significance of the regression model also need to be tested, using R^2 inspection method to test the degree of correlation between all explanatory and interpreted variables, and the significance of the model is tested by the

significance of the overall linear relationship between the explanatory and interpreted variables (Chen et al. 2017).

Beta convergence analysis. The convergence model is an important approach for analyzing regional disparities. It is based on the convergence theory of neoclassical economic growth theory (Xie and Yu, 2021). There are three types of convergence methods: sigma convergence, beta convergence, and club convergence. Among these, more scholars tend to use the beta convergence analysis method. Beta convergence is a method that links growth and its steady state. In other words, under the beta convergence, less developed regions catch up with the growth rate of developed regions over a long period. The development of primary education speed and its level of development are negatively correlated under beta convergence. Beta convergence can be divided into two types: absolute beta convergence and conditional beta convergence. Absolute beta convergence assumes that any region's primary education development level has the same "basic characteristics" (e.g., per-student education expenditure, teacher strength, level of economic development). On the other hand, conditional beta convergence does not consider the "basic characteristics" (Li and Luo, 2015).

In 1992, Barro and Sala-I-Martin proposed the classical regression equation (Barro and Sala-i-Martin, 1992) for testing absolute beta convergence:

$$\frac{1}{T} \log \left(\frac{y_{s,t+T}}{y_{s,t}} \right) = a + b \log(y_{s,t}) + \varepsilon_{s,t}$$

where y denotes the level of development of primary education, s refers to overall, urban, and rural areas, respectively, t is a point in time, T denotes the number of periods in the sample, and $\varepsilon_{s,t}$ is the random error term; in addition, $b = -\frac{1 - e^{-\beta T}}{T}$ and beta denotes the speed of convergence. When $b < 0$, the examined region shows absolute beta convergence (Han and Li, 2013) within the period T. In this study, to make the best use of the data as much as possible, let $T = 1$.

The development level of primary education may vary greatly between urban and rural areas due to their internal structural characteristics. Therefore, the absolute beta convergence analysis method needs to be modified by adding relevant control variables to establish the conditional beta convergence of the development level of primary education. This is expressed as

$$\frac{1}{T} \log \left(\frac{y_{s,t+T}}{y_{s,t}} \right) = a + b \log(y_{s,t}) + \sum_{k=1}^{Z} c_k \log(x_{k,(s,t)}) + \varepsilon_{s,t}$$

where k ($k = 1, 2, \cdots, z$) denotes the number of control variables added, and the rest of the variables have the same meaning as in the absolute beta convergence formula. As above, let $T = 1$. Since the independent variables in the model are all delayed T-periods, the probability of endogenous problems is reduced.

Empirical results
Analysis of index weights. Figure 1, derived from Table 1, shows the weights obtained using the entropy method are nearly identical to those derived from the coefficient of variation method, except for some minor differences. Consequently, the combined weights exhibit a higher degree of overlap with the results from the two methods. It shows that the methods used in this study are robust and reasonable. The weights of $x_1, x_2, x_5, x_6, x_{14}$, and x_{15} are larger and close to 0.1 in Fig. 1. Notably, four out of these six indicators pertain to educational conditions, while two relate to economic and social conditions, underscoring the significant impact of resource allocation and socioeconomic support on primary education development in China, meanwhile showing

Table 1 Evaluation index system for the development level of primary education in urban and rural areas of China.

Evaluation unit	Subsystems	Sub-indicator	Formula	Orientations	Weights		
					Coefficient of variation method	Entropy method	Combination
Level of development of primary education	Educational conditions	Per student education expenditure x_1	Funding for primary education/Average annual number of students (yuan)	+	0.116	0.104	0.112
		Per capita output value of fixed assets x_2	Gross fixed asset value of primary schools/Average annual number of students (million yuan)	+	0.103	0.084	0.094
		Floor space of school buildings per student x_3	Sum of floor space of primary school buildings/Average annual number of students (m^2)	+	0.055	0.108	0.078
		Books per student x_4	Total number of books in primary schools/Average annual number of students (books)	+	0.084	0.054	0.068
		Computers per student x_5	Total number of computers in primary schools/Average annual number of students (units)	+	0.121	0.102	0.113
		Percentage of schools with a campus network x_6	Schools with a campus network/Total number of schools (%)	+	0.117	0.112	0.116
	Teacher Strength	Ratio of teachers to students x_7	Total number of primary school teachers/Average annual number of students (%)	+	0.048	0.060	0.055
		Proportion of full-time teachers x_8	Total number of full-time primary school teachers/Total number of primary school teachers (%)	+	0.003	0.028	0.009
		Average years of education received by teachers x_9	The academic qualifications of primary school teachers are calculated (teachers' academic qualifications are classified as postgraduate, undergraduate, specialist, high school, below high school, etc., and are assigned the values of 19, 16, 15, 12, and 9 years, respectively).	+	0.012	0.016	0.014

Table 1 (continued)

Evalua tion unit	Subsys tems	Sub-indicator	Formula	Orientations	Weights		
					Coefficient of variation method	Entropy method	Combination
		Structure of teacher titles x_{10}	Calculated based on the titles of primary school teachers (the titles of teachers are classified as senior secondary school, senior primary school, primary school level 1, primary school level 2, Primary school level 3, and unassessed (after 2019, they will be classified as senior, vice-senior, intermediate, assistant, post, and unassessed), with the values assigned as 12, 10, 8, 6, 4, and 2, respectively).	+	0.010	0.009	0.009
	Student development	The number of primary schools per student x_{11}	Number of primary schools/ Average annual number of students (number)	+	0.052	0.052	0.053
		Number of classes per student x_{12}	Number of classes/Average annual number of students (number)	+	0.034	0.036	0.036
		Enrollment rate x_{13}	Number of students entering junior high school/ Number of students graduating from primary school in the current year (%)	+	0.004	0.011	0.006
	Socio- economic conditions	Per capita GDP x_{14}	Gross domestic product per capita (yuan)	+	0.115	0.107	0.112
		Per capita disposable income x_{15}	Combined average of household income (yuan)	+	0.105	0.098	0.103
		Dependency ratio x_{16}	Ratio of non-working age population to working age population (%)	−	0.021	0.019	0.020

Data for all indicators are disaggregated by overall (urban and rural), urban (aggregated for cities and towns) and rural, respectively, and in this paper overall, urban and rural primary schools refer to ordinary primary schools, excluding adult primary schools.
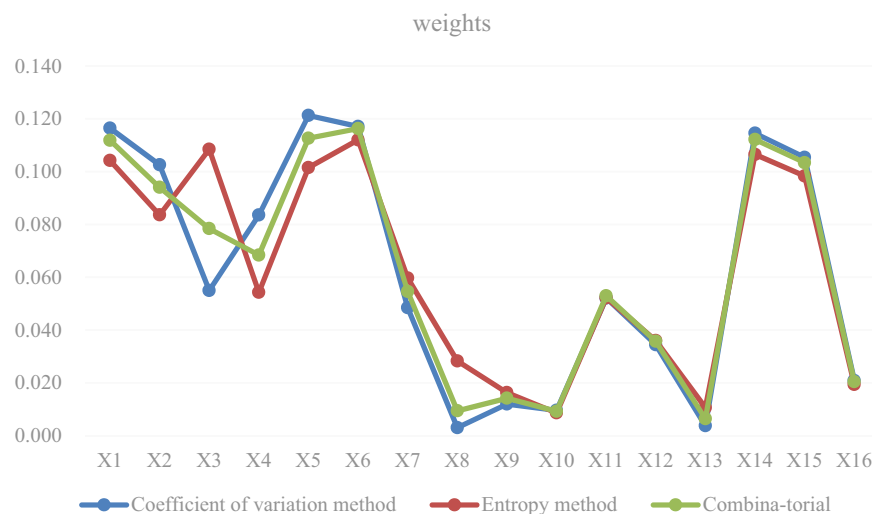
Fig. 1 Weights of the indicators of primary education level in China.

the internal data changes of these 6 indexes (i.e. the difference between urban and rural areas) are more obvious, which can more sharply reflect the operation of the evaluation system. x_8, x_9, x_{10}, and x_{13} have the lowest weights, below or around 0.01. Three of these four indicators come from teacher strength and one from student development, suggesting their relatively lower impact and smaller internal data change degree (see Appendix ①) on primary education development.

Evolution of primary education in urban and rural China throughout history. Using the entropy variation coefficient-comprehensive index method to measure the level of primary education development in China's nation, urban areas, and rural areas.

Evolutionary characterization of sub-indicators. Panel A in Fig. 2 highlights the significant improvement in educational conditions over time, with the gap between urban and rural areas initially widening before narrowing. From 1997 to 2022 (excluding 1995–1996 to avoid artificially inflated annual growth rates caused by extremely low initial values in rural areas), the average annual growth rates for educational conditions at the national, urban, and rural contexts were 13.32%, 10.51%, and 14.88%, respectively. These trends can be attributed to three main factors. First, the overall improvement in educational conditions reflects the government's prioritization of basic education and increased investment, such as the 13.62% annual growth rate in per capita education expenditure x_1. Second, rural education conditions have significantly improved due to targeted policies, including the "Two-free and one-subsidy" policy for rural compulsory education launched in 2001 (Guo et al. 2024), the "Deepening reform of the mechanism for guaranteeing rural compulsory education funding" introduced in 2005, and the "Targeted poverty alleviation policy" initiated in 2015 (see Appendix ②). Third, urban education conditions developed rapidly before 2015, leading to a widening gap in rural areas. However, since 2015, this gap has started to narrow due to the implementation of precision poverty alleviation policies, with education being a central component. During this phase, significant resources were allocated to rural areas. Despite these efforts, the gap remains substantial due to the historical under-investment in rural education compared to urban areas.

Panel B in Fig. 2, the national teacher strength demonstrates a gradual increase followed by a slight decline. The difference

between the teacher strength indices in urban and rural areas initially narrows, then widens, and finally narrows again. From 1995 to 2022, the average annual growth rates for teacher strength in national, urban, and rural contexts were 3.00%, 1.72%, and 3.35% respectively, with rural areas experiencing the fastest growth rate. Regarding the gap between urban and rural teacher strength, the difference in indices decreased from 0.26 in 1995 to −0.08 in 2001. It then increased from 0.04 in 2002 to 0.61 in 2016, before shrinking again from 0.57 in 2017 to 0.31 in 2022. As shown in Fig. 1, the ratio of teachers to students is a key influencing factor for teacher strength. The trends in teacher strength and the ratio of teachers to students exhibit significant convergence (see Appendix ③).

Panel C in Fig. 2, the student development index showed a downward trend, with a pronounced decline in both the overall and rural areas, whereas urban areas experienced two growth phases between 1995–1999 and 2009–2014. The rural student development index was larger than that of the urban areas until 2011. However, a sharp increase in the urban index post-2011, followed by the hump-shape since then, saw it almost on par with the rural index by 2022. According to Fig. 1, the weight of the number of primary schools per student and the number of classes per student was the largest in the student development subsystem, underscoring the importance of school-student concentration and small class sizes for student development (see Appendix ④). The reasons for the trend in student development in the paper are the changes in compulsory education policy, the acceleration of urbanization, and the reduction of the birth rate (Hannum and Wang, 2022; Zhang, 2017; Zhang et al. 2015) (see Appendix ⑤).

Panel D in Fig. 2, the indices for national, urban, and rural socioeconomic development have demonstrated a continuous and stable upward trend. From 1995 to 2022, the average annual growth rates were 10.54%, 8.00%, and 11.52%, respectively. The growth of these socioeconomic development indices was primarily driven by sustained increases in GDP and disposable income per capita. As illustrated in Fig. 1, GDP per capita and disposable income per capita are the most significant components of the socioeconomic development sub-system, essentially explaining the overall growth of the social economy. Additionally, the gap between urban and rural areas has been widening steadily since 1995. This is largely due to the underdeveloped state of rural socioeconomic infrastructure. Although the growth rate of the rural socioeconomic development index is higher, the gap between
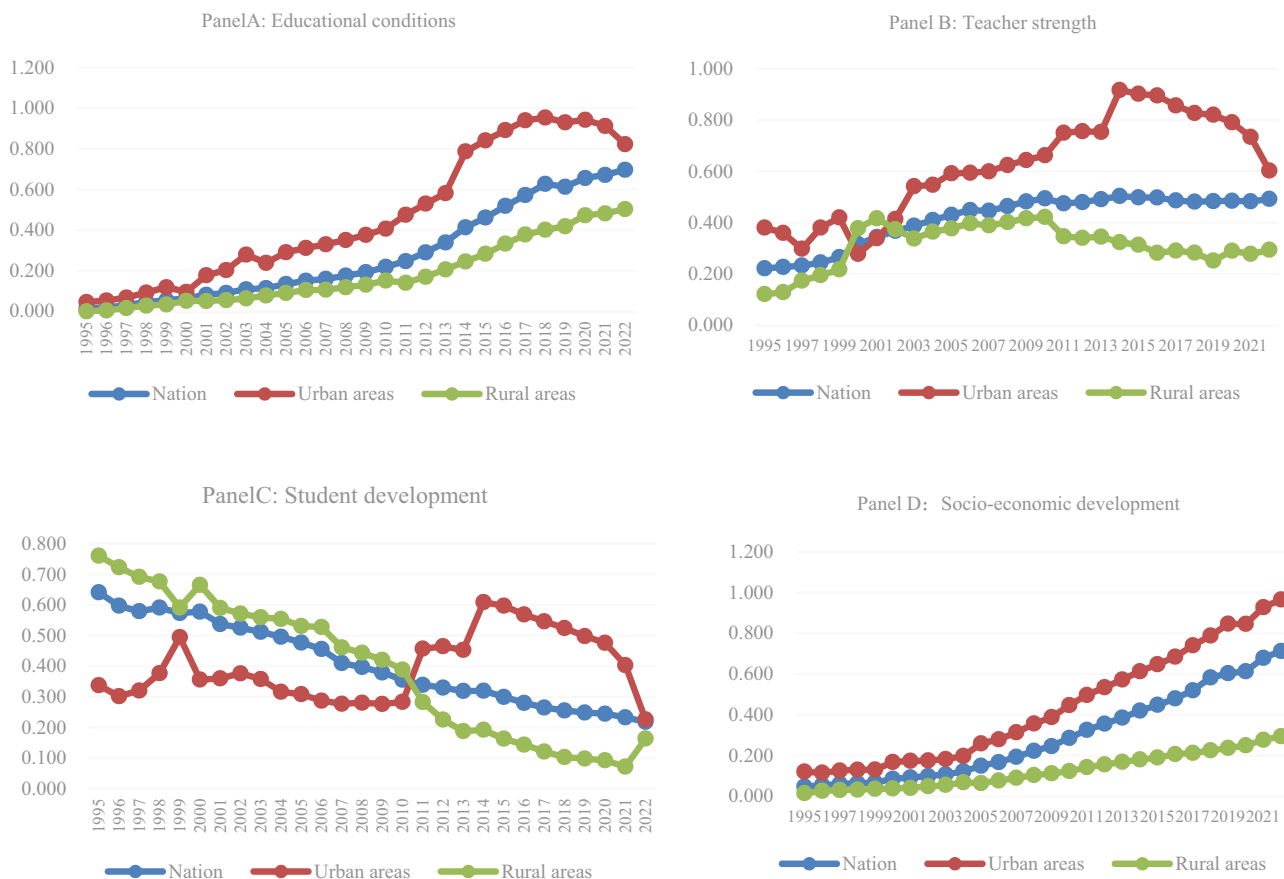
Fig. 2 Historical evolution trend of sub-system comprehensive index measurement results.
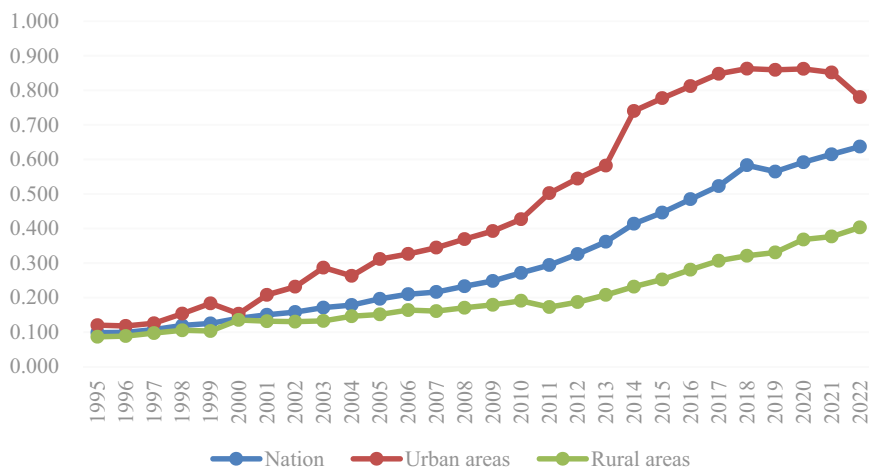

Fig. 3 Evolution of primary education level in urban and rural China.

urban and rural areas has still increased at an average annual rate of 7.09%. This indicates that without addressing the growing income disparity between urban and rural populations, it will be challenging to foster social and economic development and achieve convergence.

Analysis of general evolutionary features. Based on Fig. 3, primary education development in China has experienced a steady upward trend from 1995 to 2022. The average annual growth rates for primary education development were 6.89%, 6.91%, and 5.65%, for the national, urban, and rural areas, respectively. The higher growth rate in urban areas compared to rural areas has led to a widening gap in education development between urban and rural regions before 2018, a trend that began to shift in 2019, and the gap between urban and rural areas is narrowing (see Appendix ⑥).

Predicting the development of primary education in urban and rural areas of China

Predictor screening based on time-difference correlation analysis. The development of primary education in China is influenced by several factors. This study has identified 16 core indicators for analyzing the development level of primary education in both urban and rural areas. Using SPSS26.0 software, a time-difference correlation analysis was conducted on these indicators. The development level of primary education in overall, urban, and rural areas from 1995 to 2022 is used as the benchmark indicator. The 16 sub-indicators were taken as the selected indicators affecting the benchmark indicator. The correlation coefficients of the selected indicators relative to the benchmark indicators are calculated for several periods of overrunning, synchronousness, or lagging (typically over three years, with this study using a 5-year moving period). The resulting classification categorized the indicators as leading, synchronous, or lagging based on their time-difference correlation coefficients (absolute values exceeding 0.5).

The analysis of the time difference correlation method yields Table 2. This table includes seven priority indicators in general: x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, and x_{16}. Additionally, there are seven priority indicators for urban areas and six priority indicators for rural areas. The priority indicators for the level of development of primary education in urban and rural areas are partially consistent with the whole. However, urban areas have three additional indicators, including x_3, x_4, and x_{13}, rural areas have two additional indicators, including x_1 and x_{13}. The principle of time difference correlation analysis considers synchronous and lagging indicators, as well as indicators with a time difference correlation coefficient with an absolute value of <0.5, to have a weaker influence on the benchmark indicators. On the other hand, the leading indicators with a correlation coefficient with an absolute value of more than 0.5 can better reflect and predict the benchmark indicators.

All prior indicators of the development level of primary education in overall and urban areas have correlation coefficients >0.5. Therefore, they are considered as main factors affecting the development level of primary education. For rural areas, only one leading indicator is considered a secondary factor because its absolute correlation coefficient is <0.5. The excluded metric is x_{13}.

Prediction and testing of multiple linear regression models.

(1) Model construction

Using the time difference correlation analysis method, the study identified the early warning factors that affect the development levels of primary education in overall, urban, and rural areas. The dependent variable, in this case, was the development level of primary education. Regression analysis was performed using Eviews 8.0 software to obtain the multivariate regression prediction model for primary education development levels in China's overall, urban, and rural areas:

$$y_{\text{nation}} = -0.905 - 1.397 \times 10^{-2} x_7 - 2.407 \times 10^{-3} x_8$$
$$+ 1.332 \times 10^{-1} x_9 - 1.056 \times 10^{-1} x_{10} - 54.196 x_{11}$$
$$+ 17.409 x_{12} + 3.679 \times 10^{-3} x_{16}$$

$$y_{\text{urban areas}} = 0.570 + 1.359 \times 10^{-1} x_3 + 2.790 \times 10^{-5} x_4$$
$$- 3.679 \times 10^{-2} x_7 + 6.340 \times 10^{-4} x_{10} - 15.917 x_{12}$$
$$- 3.745 \times 10^{-3} x_{13} - 1.708 \times 10^{-3} x_{16}$$

$$y_{\text{rural areas}} = -6.527 \times 10^{-2} + 2.650 \times 10^{-5} x_1 + 6.296 \times 10^{-3} x_7$$
$$- 1.640 \times 10^{-3} x_8 - 60.977 x_{11} + 18.848 x_{12}$$

(2) Validity test and modification of the model

In general, Pearson correlation analysis requires that the absolute value of the correlation coefficient between the two variables is >0.5, and the corresponding *P*-values is <0.1. The Multicollinearity test for variance inflation factor (VIF) requires IF values of <10 for the independent variables. The closer the value of R^2 is to 1, the better the fitting degree of the model is. The larger the *F*-statistic and the smaller the *P*-value of the model, the more significant the effect of the independent variables on the dependent variable. According to the above test method, it is necessary to eliminate x_7, x_8 and x_{11} in the prediction model of the overall primary education development level, while urban areas should discard x_4, x_{10}, x_{13} and x_{16}, and rural areas should get rid of x_7 and x_8. Thus, the appropriate and effective modified prediction models are obtained.

Modified multiple regression prediction model for the primary education development levels in China's overall, urban, and rural areas:

$$y_{\text{whole amendments}} = -1.706 + 0.174 x_9 - 0.125 x_{10} + 13.809 x_{12}$$
$$+ 4.171 \times 10^{-3} x_{16}$$

$$y_{\text{urban amendments}} = 0.123 + 0.130 x_3 - 3.292 \times 10^{-2} x_7 - 14.600 x_{12}$$

$$y_{\text{rural amendment}} = -0.211 + 2.62 \times 10^{-5} x_1 - 62.013 x_{11} + 19.838 x_{12}$$

(3) Determination of the reasonableness of the model

The data from 1995 to 2022 for overall, urban and rural areas in China were used to predict the level of development of primary education. The predicted value was then compared with the actual value for each year, and the relative errors were analyzed. The results are presented in Table 3. The table shows that the absolute relative errors of overall, urban and rural areas in the table are >10% for 2, 5, and 3 years, respectively, and <10% for all other years. Therefore, the forecasts are more accurate except for a few years, it can be concluded that the model used to predict the level of development of primary education is reasonable.

To further assess the model's validity in predicting the development level of primary education, data from 1995 to 2017 were used to project development levels for 2018–2022 in national, urban, and rural areas in China. The data of early warning indicators derived from the 2018 to 2022 prediction were applied to a multivariate linear regression model to obtain the predicted value of the primary education development levels during this period. Then, compare the predicted value with the actual value and according to the relative error judgment the accuracy of the model's prediction of the dependent variable based on the estimated independent variable.

The revised prediction model incorporates four indicators in nation: x_9, x_{10}, x_{12}, and x_{16}. The average years of teacher education and the structure of teacher titles have been increasing year by year, with an average annual growth rate of 1% and 0.5%, respectively, use those as the projection estimate for x_9 and x_{10} for the next 5 years. In response to the problem of ageing, China has changed its birth policy to allow three children since 2021, but the effect of the policy has yet to be seen, with negative population growth for two consecutive years so far. Considering the possible lag period of the policy support effect, a negative growth rate of 3% is estimated for x_{12} in the next 5 years based on historical data with an average annual decline rate of 4.8%. The dependency ratio had declined by an average of 2% annually until 2010 but had risen by 2% since 2011, taking into account the impact of the previous family planning policy, and a 2% increase was used as a projection estimate for x_{16} in the next 5 years.

For urban primary education, the revised prediction model includes three indicators: x_3, x_7, and x_{12}. x_3 uses the projected

Table 2 Leading, synchronous, and lagging indicators for the development level of primary education.

Indicators	Nation			Urban areas			Rural areas		
	Type of indicators	Pilot strength	Pilot length	Type of indicators	Pilot strength	Pilot length	Type of indicators	Pilot strength	Pilot length
Per student education expenditure x_1	Synchronous	0.994	0	Lagging	0.989	1	Leading	0.990	−4
Per capita output value of fixed assets x_2	Synchronous	0.987	0	Synchronous	0.977	0	Synchronous	0.984	0
Floor space of school buildings per student x_3	Synchronous	0.980	0	Leading	0.985	−1	Synchronous	0.299	0
Books per student x_4	Synchronous	0.936	0	Leading	0.976	−1	Synchronous	0.538	0
Computers per student x_5	Synchronous	0.996	0	Synchronous	0.986	0	Synchronous	0.993	0
Percentage of schools with a campus network x_6	Synchronous	0.989	0	Synchronous	0.960	0	Synchronous	0.977	0
Ratio of teachers to students x_7	Leading	0.672	−4	Leading	0.930	−2	Leading	−0.771	−1
Proportion of full-time teachers x_8	Leading	0.965	−2	Lagging	0.924	5	Leading	0.663	−4
Average years of education Received by teachers x_9	Leading	0.926	−1	Synchronous	0.792	0	Synchronous	0.719	0
Structure of teacher titles x_{10}	Leading	0.648	−5	Leading	0.615	−5	Lagging	−0.565	5
The number of primary schools per student x_{11}	Leading	−0.945	−1	Synchronous	−0.714	0	Leading	−0.931	−2
Number of classes per student x_{12}	Leading	−0.722	−4	Leading	−0.578	−5	Leading	−0.942	−2
Enrollment rate x_{13}	Synchronous	0.649	0	Leading	0.679	−1	Leading	−0.423	−2
Per capita GDP x_{14}	Synchronous	0.992	0	Lagging	0.976	1	Synchronous	0.980	0
Per capita disposable income x_{15}	Synchronous	0.995	0	Lagging	0.981	1	Synchronous	0.988	0
Dependency ratio x_{16}	Leading	−0.733	−5	Leading	0.729	−5	Lagging	0.685	5

Table 3 Relative error between predicted and actual values of primary education development level in China from 1995 to 2022.

Vintages	Nation			Urban areas			Rural areas		
	Actual value	Projected value	Relative error (%)	Actual value	Projected value	Relative error (%)	Actual value	Projected value	Relative error (%)
1995	0.099	0.111	−12.554	0.120	0.076	−36.952	0.087	0.087	0.730
1996	0.099	0.101	−1.242	0.118	0.131	11.069	0.088	0.097	9.985
1997	0.108	0.102	5.629	0.126	0.148	17.666	0.097	0.104	6.795
1998	0.119	0.119	0.349	0.153	0.134	−12.510	0.105	0.109	3.899
1999	0.125	0.124	1.253	0.183	0.201	9.463	0.103	0.089	−13.740
2000	0.140	0.120	14.378	0.153	0.161	5.260	0.135	0.126	−6.948
2001	0.150	0.141	6.220	0.208	0.198	−4.732	0.132	0.129	−2.126
2002	0.158	0.151	4.191	0.231	0.243	4.990	0.130	0.131	0.718
2003	0.171	0.169	0.833	0.287	0.281	−2.081	0.132	0.130	−1.422
2004	0.178	0.193	−7.958	0.263	0.300	14.112	0.146	0.145	−0.911
2005	0.196	0.209	−6.236	0.311	0.330	5.908	0.151	0.144	−4.620
2006	0.210	0.222	−5.892	0.326	0.348	6.647	0.164	0.153	−6.796
2007	0.216	0.218	−1.050	0.344	0.363	5.336	0.161	0.148	−8.147
2008	0.233	0.237	−1.649	0.369	0.361	−2.390	0.171	0.163	−4.344
2009	0.248	0.258	−4.043	0.393	0.381	−2.963	0.179	0.184	2.803
2010	0.271	0.269	0.854	0.427	0.417	−2.300	0.191	0.203	6.464
2011	0.294	0.296	−0.812	0.502	0.455	−9.395	0.172	0.176	1.888
2012	0.326	0.328	−0.692	0.544	0.514	−5.546	0.187	0.210	12.697
2013	0.362	0.358	1.012	0.582	0.586	0.740	0.208	0.235	13.054
2014	0.414	0.404	2.406	0.740	0.757	2.244	0.232	0.238	2.612
2015	0.446	0.433	2.857	0.777	0.792	1.894	0.252	0.266	5.247
2016	0.485	0.463	4.572	0.812	0.823	1.358	0.281	0.293	4.239
2017	0.523	0.537	−2.771	0.848	0.858	1.193	0.307	0.307	0.168
2018	0.583	0.579	0.709	0.863	0.865	0.251	0.321	0.316	−1.475
2019	0.565	0.582	−3.093	0.859	0.859	−0.065	0.331	0.327	−0.985
2020	0.592	0.611	−3.265	0.862	0.869	0.802	0.368	0.338	−8.147
2021	0.615	0.620	−0.836	0.852	0.804	−5.565	0.377	0.345	−8.313
2022	0.637	0.608	4.580	0.780	0.779	−0.201	0.403	0.423	4.903

According to Liu and Fang (2017), the classification criteria of prediction accuracy grade are: |relative error|<3.5% is good; 3.5%≤|relative error|<5% is qualified; 5%≤|relative error|<10% is barely qualified; |relative error|≥10% is unqualified.

Table 4 Prediction of the development level of primary education from 2018 to 2022.

Vintages	Nation			Urban areas			Rural areas		
	Actual value	Projected value	Relative error (%)	Actual value	Projected value	Relative error (%)	Actual value	Projected value	Relative error (%)
2018	0.583	0.551	−5.512	0.863	0.803	−6.957	0.321	0.324	1.070
2019	0.565	0.565	0.116	0.859	0.824	−4.087	0.331	0.344	3.973
2020	0.592	0.581	−1.930	0.862	0.845	−1.962	0.368	0.365	−0.736
2021	0.615	0.596	−3.032	0.852	0.865	1.601	0.377	0.389	3.347
2022	0.637	0.613	−3.857	0.780	0.885	13.434	0.403	0.416	3.148

growth rate of 3%, consistent with the national projection. x_7 was calculated based on the average of the past 6 years, while x_{12} was predicted using the 1995–2017 average for the next 5 years.

The revised rural primary education prediction model comprises three indicators: x_1, x_{11}, and x_{12}. Since sustainable macroeconomic growth rates typically range from 2% to 6%, the average annual growth rates of 8% are used to estimate x_1 for 2018–2022. Taking into account state preferential policy support for primary education, the same negative growth rate of 3% is used as the predicted values for the next 5 years for x_{11} and x_{12}.

This predictive approach was used to estimate the development level of primary education across China's overall, urban, and rural areas. The estimated values were compared against actual data to assess the model's accuracy and validity. The results, summarized in Table 4, indicate that the relative error for urban primary school development in 2022 was "unqualified", while the rest predictions were rated as "Good", "Qualified" or "Barely

qualified". Moreover, only the relative error of the national and urban predicted values in 2018 was rated as "Barely qualified". The values of the indicators estimated by this method are feasible, and the model's predictive results were found to be reasonable and valid.

(4)   Forecast results and analyses

The following indicators were estimated while considering current trends and conditions. For national primary education development, the average years of teacher education were projected with a 0.3% annual increase over the last 5 years, and the structure of teacher titles was estimated on the 1995–2022 average. As the impact of the birth policy has gradually become apparent, the average number of classes per student has changed little in the last 5 years, with a decline rate of <1%. Therefore, it is projected that the average number of classes per student will decline by 0.5% for 2023–2035. Due to the increasing challenges
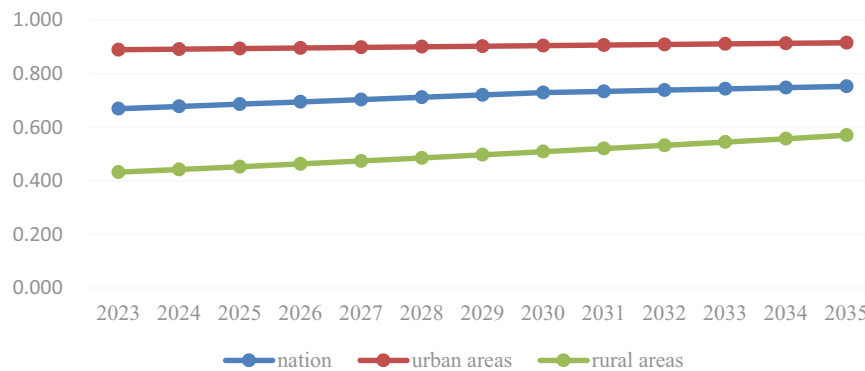
Fig. 4 Prediction of the development level of primary education in urban and rural areas of China from 2023 to 2035.
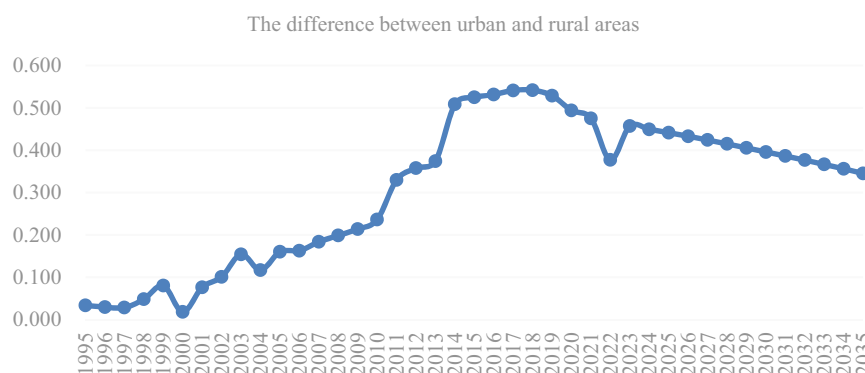

Fig. 5 The development level gap of primary education in urban and rural areas of China from 1995 to 2035.

of an aging population, China's dependency ratio has been rising continuously since 2010, with an average annual increase rate of 2%. Therefore, the dependency ratio in the context of the fertility policy change is estimated at an increase of 1% from 2023 to 2030 and a decrease of 1% from 2031 to 2035.

For estimating the development of primary education in urban areas, the floor space of school buildings per student was estimated at an average of 11.5 square meters over the past 15 years, and the ratio of teachers to students was estimated at 8.5% of the average from 1995 to 2022, and the estimated average number of classes per student is the same as the national forecast rate of decline by 0.5%.

In estimating the level of development of rural primary education, the median of the sustainable growth rate of macro-economy is 3%, which is the forecast of the average annual growth rate of the per student education expenditure in the future. And the rates of change for the number of primary schools per student and the number of classes per student were projected to follow the same trend as the national rate of change for the number of classes per student.

These estimated values were applied to the model to forecast the development levels of primary education in national, urban and rural areas from 2023 to 2035. The data presented in Fig. 4 indicates that the primary education development level in China is gradually improving, but at a slow pace across all categories, including national totals, urban areas, and rural areas (China's economic development level has changed from high speed to steady growth, and the investment forecast of primary education development level is consistent with it, so the forecast chart Fig. 4 is no longer as steep as the actual chart Fig. 3 in the past). Additionally, while the gap between rural and urban primary

education levels in China has narrowed, the reduction remains modest, highlighting that significant imbalances in urban and rural primary education development persist.

Figure 5 illustrates the disparity in the development levels of primary education between urban and rural areas, as derived from Figs. 2 and 4. From 1995 to 2018, this gap continued to widen. However, between 2019 and 2035, the gap gradually narrowed due to significant state support for rural education infrastructure, teacher recruitment, student development, and other key areas, resulting in an annual average reduction rate of 2.63%. Although the gap in primary education levels between urban and rural areas in China has decreased, this decline is still modest, highlighting the persistent imbalance in the development of primary education across the two regions.

Convergence analysis of the level of development of primary education in urban and rural areas in China. As can be seen from Table 5, the regression coefficients for the overall, urban and rural primary education development levels are negative, whether we consider the actual data of 1995–2022 or include the predicted data of 2023–2035. In addition, only from 1995 to 2035, the development level of primary education in nation and urban areas has a significant absolute convergence, the rest is not significant absolute convergence. In terms of the remarkable absolute convergence, the convergence rate of the general and urban primary education development level was slower, and the urban convergence rate was faster than that of the general from 1995 to 2035, therefore, it implies that the development level of primary education in rural areas is extremely unbalanced and difficult to converge.

Table 5 Absolute beta convergence of the development level of primary education.

Vintages		1995–2022							
		Coefficient	Std. error	t-Statistic	P-value	R^2	Adjusted R-squared	F-statistic	Convergence rate
Nation	x	−0.006	0.012	−0.538	0.595	0.011	−0.028	0.289	–
	C	0.060***	0.018	3.391	0.002				
Urban areas	x	−0.046	0.030	−1.552	0.133	0.088	0.051	2.409	–
	C	0.023	0.036	0.653	0.519				
Rural areas	x	−0.073	1.307	−0.056	0.956	0.000	−0.040	0.003	–
	C	−1.734***	0.113	−15.290	0.000				

Vintages		1995–2035							
Type	Explanatory variables	Coefficient	Std. error	t-Statistic	P-value	R^2	Adjusted R-squared	F-statistic	Convergence rate
Nation	x	−0.031***	0.008	−3.913	0.000	0.287	0.268	15.314	0.014
	C	0.018*	0.010	1.871	0.069				
Urban areas	x	−0.054***	0.020	−2.709	0.010	0.162	0.140	7.341	0.024
	C	0.013	0.020	0.639	0.527				
Rural areas	x	−2.269	1.673	−1.356	0.183	0.046	0.021	1.838	–
	C	−1.303***	0.123	−10.627	0.000				

"***", and "*" in the upper right corner of the coefficients in the table represent significance at the 1% and 10% levels, respectively, and "x" is the level of development of primary education.

Table 6 Beta convergence of conditions for the development level of primary education.

Vintages					1995–2022				
		Coefficient	Std. error	T-statistic	P-value	R^2	Adjusted R-squared	F-statistic	Convergence rate
Nation	x	0.269	0.180	1.497	0.149	0.332	0.172	2.083	–
	x_9	−2.401	1.422	−1.688	0.106				
	x_{10}	0.799	0.505	1.582	0.129				
	x_{12}	−0.833*	0.422	−1.972	0.062				
	x_{16}	0.002	0.155	0.011	0.992				
	C	2.137	2.341	0.913	0.372				
Urban areas	x	−0.170	0.341	−0.498	0.624	0.148	−0.007	0.954	–
	x_3	0.299	0.848	0.352	0.728				
	x_7	0.101	0.252	0.402	0.692				
	x_{12}	−0.315	0.493	−0.639	0.529				
	C	−1.990	3.286	−0.605	0.551				
Rural areas	x	−0.266	0.184	−1.447	0.162	0.367	0.252	3.190	–
	x_1	−0.014	0.037	−0.373	0.712				
	x_{11}	−0.020	0.419	−0.047	0.963				
	x_{12}	−0.431	0.531	−0.813	0.425				
	C	−2.032*	1.026	−1.980	0.060				

Vintages					1995–2035				
Type	Explanatory variables	Coefficient	Sth. error	T-statistic	P-value	R^2	Adjusted R-squared	F-statistic	Convergence rate
Nation	x	0.242*	0.125	1.936	0.061	0.623	0.567	11.217	–
	x_9	−2.174**	0.998	−2.179	0.036				
	x_{10}	0.737**	0.335	2.202	0.035				
	x_{12}	−0.575*	0.288	−1.996	0.054				
	x_{16}	−0.084	0.103	−0.807	0.425				
	C	2.862*	1.631	1.755	0.088				
Urban areas	x	−0.059	0.270	−0.219	0.828	0.198	0.107	2.163	–
	x_3	−0.056	0.586	−0.095	0.925				
	x_7	0.165	0.148	1.113	0.273				
	x_{12}	−0.067	0.243	−0.275	0.785				
	C	−0.444	2.145	−0.207	0.837				
Rural areas	x	−0.002	0.074	−0.031	0.975	0.253	0.167	2.957	–
	x_1	−0.012	0.033	−0.367	0.716				
	x_{11}	0.264	0.237	1.114	0.273				
	x_{12}	−0.543	0.338	−1.607	0.117				
	C	−0.358	0.361	−0.992	0.328				

The "**" and "*" in the upper right corner of the coefficients in the table represent significance at the 5% and 10% levels, respectively.

Combining the main factors affecting the development level of primary education in national totals, urban, and rural areas in the aforementioned multiple regression linear prediction model, the eight control variables of x_1, x_3, x_7, x_9, x_{10}, x_{11}, x_{12}, and x_{16} are added to the conditional beta convergence to obtain the analyses in Table 6. According to the actual data from 1995 to 2022, there is a not significant conditional beta convergence in the level of development of primary education in urban and rural areas, in which all the factors influencing the level of development of primary education are insignificant. Together with the predicted data, there is conditional beta convergence for both rural and urban primary education development levels, but they are not significantly converging, and all the influencing factors are indistinctive.

Conclusion and discussion

The purpose of this study is to establish a comprehensive evaluation system for assessing the level of development of primary education. The research utilized national, urban and rural areas from 1995 to 2022 of time series data (full sample data at national level). The entropy variation coefficient-comprehensive index method is applied to explore the historical trends in the developments of primary education in China's urban and rural regions. The concept of early warning was integrated through time-difference correlation analysis to identify key factors influencing primary education development. These factors are then used as explanatory variables to construct a multivariate regression prediction model, which anticipates the future development trend of primary education from 2023 to 2035. Finally, the research assessed the convergence of the level of development of primary education using beta convergence analysis. The main findings of this study can be summarized in four key points.

Firstly, the study found that primary education in China's urban and rural areas has shown steady improvement (see Fig. 3). However, due to the faster growth rate in urban areas compared to rural areas, the urban-rural divide was still significant. Currently, there are only few studies that have measured the development level of urban and rural education in China. It is generally believed that the gap between urban and rural education in China has been narrowing, driven by national policy shifts, rapid industrialization, and other factors (Luo et al. 2022; Guo and Li, 2024). Two main perspectives emerge from existing research: one emphasizes the heterogeneity in regional education investment, teacher quality, and student demographics, leading to a widening gap in education development (Tu et al. 2005; Cai and Song, 2021). The other suggests that factors such as large-scale migration driven by industrialization, regional coordinated development strategies, reform of the household registration system, and other factors have gradually reduced the degree of imbalance in China's regional education development (Xu and Liu, 2023; Xie and Yu, 2021). Our conclusion diverge from most studies because most studies used single indicators such as school enrollment, years of education, and academic achievement. For example, Zhang et al. (2015) focused on academic achievement as dependent variable, and uses data from the China Family Panel Survey (CFPS) and the Rural–Urban Migration in China (RUMiC) survey to compare the education performance of rural children, children of rural-to-urban migrants, and urban children over the period 2009–2010. They find that there is still significant inequality in the education performance of children between rural and urban areas of the PRC. In particular, rural-to-urban migration, which had been expected to play an active role in reducing education inequality, could not contribute much to narrow the gap. Similarly, Zhang (2017) also uses the Rural–Urban Migration in China (RUMiC) survey, takes school enrollment as dependent variable, and applies logistic regression and survival

analysis to illustrate the new education-poverty trap imposed on migrant children by the institutional constraints and hierarchies in children's education, created by the Chinese household registration system in Chinese cities. While we take the development level of primary education as the dependent variable, which is a comprehensive evaluation system that incorporates both internal and external factors. We have deconstructed the level of education development in urban and rural areas into four aspects: educational conditions, teacher strength, student development, and socioeconomic development. Four indicators show that education in urban areas is better than in rural areas (see Fig. 2). The combination of these sub-indicators reveals a widening gap between urban and rural primary education development levels (it started to shrink in 2022). The study highlights obstacles hindering the convergence of primary education development in China's urban and rural areas, including educational infrastructure, teacher competency, student outcomes and socioeconomic factors. Rural areas face a weaker foundation, so priority should be given to improving educational infrastructure, teacher competency, student development and supporting socioeconomic growth. The government should thus focus on enhancing rural education, increasing teacher availability, enhancing students' academic performance, and promoting rapid socioeconomic development. These efforts are essential for balancing primary education between urban and rural areas.

Secondly, according to our projected values for 2023–2035 indicate that while the gap between urban and rural education is slowly narrowing. Our prediction results show that the level of education development in national, urban and rural areas will steadily increase from 2023 to 2035. In addition, the average annual growth rate of primary education development in urban areas is lower than that in rural areas. The gap between the two is narrowing. The existing research mainly focuses on the prediction of students' academic achievement, the demand for educational resources, and the change of the school-age population. It is evident that the existing predictions mostly use a single indicator. However, we use multiple indicators to make a useful supplement to the existing literature (Xu and Liu, 2023; Li et al. 2021; Zhou, 2024). In addition, our conclusions are based on the calculation of future investment, considering the current investment scale and its growth forecast. This finding has important implications, suggesting that the government should focus on optimizing top-level design and addressing the constraints of weak rural primary education foundations. Such constraints have led to an inability to self-regulate and catch up with urban education for a long time. These factors include education conditions, socioeconomic conditions, teacher strength and student development within the subsystem, such as education funding, number of students on campus, the average years of education received by teachers, and more.

Thirdly, our research uncovered concerning evidence that the urban–rural education divide has only slightly narrowed. The results of absolute beta convergence analysis align with our findings, showing no significant absolute convergence for national, urban and rural areas between 1995 and 2022. Although slow and significant absolute convergence with national and urban areas between 1995 and 2035, the same cannot be said for rural areas. In addition, the speed of convergence in urban areas is faster than that in general, which further indicates that the disequilibrium within rural areas is also difficult to reconcile. Our findings make a valuable contribution to the existing literature on educational convergence, highlighting the urgent need for action to address this growing inequality and ensure equal access to education for all. While previous studies noted convergence trends in national and urban education (Liaskos and Papadas, 2010; Han et al. 2013). However, our research findings are clear: continuing to invest in education at current levels in both urban

and rural areas will not sufficiently improve the differentiation of primary education development between these two areas. It is crucial to recognize that even if there is a slow convergence of primary education development within urban areas, there will still be long-term imbalances that need to be addressed. Therefore, it is essential to take action and implement targeted measures to address this issue and ensure equal access to quality education for all.

Lastly, regarding the development of primary education in urban and rural areas, conditional convergence is a key factor. However, from 1995 to 2022, there was no conditional convergence in overall, urban and rural areas, while overall and urban areas showed unremarkable conditional convergence from 2001 to 2035. After controlling for variables, it was observed that optimizing investment resources and promote the gathering of schools and students for primary education can promote a balanced development process of primary education across the nation, including both urban and rural areas. This view is supported by current research such as Sab et al. (2001) study on education and health human capital convergence, Stamatakis, Petrakis (2006) research on human capital convergence within countries, as well as Zhou et al. (2019) on educational convergence in Chinese. The study also highlights key factors influencing primary education development (see Fig. 1). These include average per-student educational expenditure, per capita fixed assets output, the percentage of schools with a campus network, computers per student, the teachers-students ratio, average years teacher education, and per capita disposable income. Enhancing teacher quality and availability, as well as increasing education investment and household income, can positively impact student numbers and overall educational development. Concentrating educational resources by encouraging student clustering can further support balanced development. Achieving balanced primary education development in urban and rural areas requires promoting income and resource balance between urban and rural residents. While some existing studies have reached similar conclusions, their focus has often been limited to specific indicators, such as graduation rates, teacher-to-student ratios, investment in education, and average years of schooling (Li et al. 2015). This lack of comprehensiveness means that a more comprehensive approach is necessary to fully understand the factors that contribute to the development of primary education.

Limitations and future research directions

It is important to note that the accuracy of our study's findings and results is limited by the reliability of our forecasts. Since our projections are based on historical data and assumption that future trends will mirror past pattern, any deviation from this assumption may result in biased forecasts and inaccuracies. To ensure the validity of our conclusions, follow-up research is recommended to validate our findings. For broader applicability, we suggest applying the research methods from this study to analyze the development of primary education in other countries, particularly developing nations with significant educational disparities. In addition, our analysis is based on national-level time series data, which limits the granularity of insights. Unlike studies such as Zhang et al. (2015) and Zhang (2017), which examine the nuanced impact of rural-urban migration and the household registration system on the development of primary education, our approach does not capture micro-level influences. Therefore, future research should also focus on understanding the micro-level factors such as policy effects, students' mental and physical health, home-school relationships, and compare the results with other types of education, such as vocational and higher education.

Data availability

The data used in this study is not accessible due to copyright regulations and our commitment to protecting the rights and information security of our data providers. The datasets utilized were primarily taken from four statistical yearbooks published by the Chinese government: *China Education Statistics Yearbook, China Educational Finance Statistics Yearbook, China Statistical Yearbook, and China Population and Employment Statistics Yearbook,* covering the years from 1995 to 2023. Additionally, a variety of other datasets can be found on the homepage of the Ministry of Education of the People's Republic of China. This resource offers annual education statistical data, educational development statistical bulletins, and important announcements regarding education funding execution at http://www.moe.gov.cn/. All data were purchased through dedicated project funding. For those interested in using this data, we recommend that researchers independently obtain these materials through the official channels mentioned above.

References

Assaad R, Krafft C (2015) Is free basic education in Egypt a reality or a myth? Int J Educ Dev 45:16–30. https://doi.org/10.1016/j.ijedudev.2015.09.001

Barro RJ, Sala-i-Martin X (1992) Public finance in models of economic growth. Rev Econ Stud 59(4):645–661. https://doi.org/10.2307/2297991

Cai WB, Cao YJ, Zhao ZQ (2022) Measurement and evaluation of the development level of education modernization in ethnic regions: based on the 2011–2019 education development index analysis. Qinghai J Ethnol 33(3):186–193

Cai WB, Song DJ (2021) Measurement and analysis of spatial characteristics of primary school education resource allocation in counties in China. Educ Teach Res 35(5):55–65

Cai YQ, Zhang CY (2023) Measurement, regional differences and dynamic evolution of preschool education development level in China. Educ Econ 39(6):11–20+30

Cai ZY, Liu ZX, Zuo SM et al. (2019) Finding a peaceful road to urbanization in China. Land Use Policy 83:560–563. https://doi.org/10.1016/j.landusepol.2019.02.042

Cao SX, Li Q (2018) Educational development priorities in three developing countries: a comparative study. Renmin Univ China Educ J 04:5–22

Chen W, Chan A (2017) Research on water resources bearing capacity of Wuhan based on multivariate linear regression model. J Henan Polytech Univ (Nat Sci) 36(01):75–79

Chu HQ (2009) Integrating urban and rural education: system reconstruction and institutional innovation: the dualistic educational structure in China and its declassification. Educ Res 30(11):3-10+26

Ding JJ (2014) Comparative analysis on poverty degree of China's 11 contiguous destitute areas: with view of comprehensive development index. Sci Geogr Sin 34(12):1418–1427

European Commission (2006) Efficiency and equity in European education and training systems. European Commission

Feng L, Zhang YY (2018) Spatial effect of education: match or mismatch of primary school? A case study of Xi' an. J Arid Land Resour Environ 32(5):19–26. https://doi.org/10.13448/j.cnki.jalre.2018.136

Field AP, Evans D, Bloniewski T et al. (2019) Predicting maths anxiety from mathematical achievement across the transition from primary to secondary education. R Soc Open Sci 6(11):191459. https://doi.org/10.1098/rsos.191459

Guo N, Wang S, Zhang J (2024) The short-and long-run impacts of free education on schooling: direct effects and intra-household spillovers. Econ J 134(663):2876–2911. https://doi.org/10.1093/ej/ueae049

Guo Y, Li X (2024) Regional inequality in China's educational development: an urban-rural comparison. Heliyon 10(4):e26249

Han HB, Li QS (2013) The convergence analysis of rural education in China: based on panel data at province level. Educ Econ 2:56–61

Hannum E, Wang F (2022) Fewer, better pathways for all? Inter-sectional impacts of rural school consolidation in China's minority regions. World Dev 151:105734

Hu SX, Dai JW, Wen HR et al. (2023) Digital transformation promotes the high-quality development of rural education: value, dilemma and strategy. J Soc Sci Humanit Lit 6(6):211–218. https://doi.org/10.53469/jsshl.2023.06(06).32

Jia X (2023) Using the hukou system to alleviate inequity in education of China. Front Educ Res 6(15). https://doi.org/10.25236/FER.2023.061503

Krueger AB, Lindahl M (2001) Education for growth: why and for whom? J Econ Lit 39:1101–1136

Lei WH, Xie XQ (2023) The third way of urban and rural compulsory education integration under the background of urbanization. Res Educ Dev 43(12):48–56. https://doi.org/10.14121/j.cnki.1008-3855.2023.12.007

Li DX, Lei F (2021) Research on the evaluation and equilibrium strategy of the allocation level of primary school education resources in China: analysis based on PLS structural equation model. Educ Theory Pract 41(19):21–26

Li K, Luo D (2015) Policy suggestions for the balanced development of compulsory education: based on the convergence analysis of 31 provinces' panel data in China. Res Educ Dev 35(10):7–14. https://doi.org/10.14121/j.cnki.1008-3855.2015.10.003

Li L, Zhu HX, Chen XL (2019) On the evaluation of human capital development in compulsory education: a perspective of the anti-poverty theory. Educ Res 40(10):124–131

Li YQ (2019) Based on empirical analysis of Yunnan's Tibetan education blocking inter-generational transfer of poverty: taking Deqin County as an example. Chin J Agric Resour Reg Plan 40(4):153–160

Liaskos G, Papadas CT (2010) Human capital convergence in Greece: a panel data analysis. Res Top Agric Appl Econ 1:141–165

Lin YF (2008) Economic development and transformation: trends, strategies, and self-sufficiency. Peking University Press

Liu CL, Fang HY (2017) Prediction of water consumption based on time difference correlation analysis and regression model. Water Sav Irrig 10:70–73+83

Liu K, Liu BT (2017) Spatial concentration tendency and unbalanced development of elementary and middle schools under the background of urbanization in central rural area: a case study of Biyang county in Henan province. Geogr Res 36(09):1680–1696

Lu HL, Zhang LG, Gao H et al. (2023) Dilemma and resolution of high-quality and balanced development of rural compulsory education from the perspective of urban and rural education integration. Theory Pract Educ 43(28):26–33

Luo G, Zeng S, Baležentis T (2022) Multidimensional measurement and comparison of China's educational inequality. Soc Indic Res 163(2):857–874

Ma JL, Chen XY, Wei H (2020) The spatial layout and formation mechanism of primary education resources in China. Educ Acad Mon 10:14–22

Meng WQ, Yao Y (2022) China's path to high-quality and balanced development of compulsory education from an international perspective. Educ Res 43(06):83–98. http://www.moe.gov.cn/

Mo DX, Han XP, Ying H (2022) The changes and differences in the level of development of pre-school education between provinces in China. Mod Educ Manag 11:98–109

Molina-Muñoz D, Contreras-García JM, Lara-Porras AM et al. (2023) Determinants of the Spanish students' performance in mathematics in PISA: a multilevel analysis. Educ Res Eval 28(4–6):116–129. https://doi.org/10.1080/13803611.2023.2276194

Nascimento RLS, Fagundes RAA, de Souza RMCR (2022) Statistical learning for predicting school dropout in elementary education: a comparative study. Ann Data Sci 9(4):801–828. https://doi.org/10.1007/s40745-021-00321-4

North DC Institutions (1990) Institutional change and economic performance. Cambridge University Press

Ojha AK, Chakravarty D (2012) Beyond corporate social responsibility: a role for corporate India in rural primary education. IIM Kozhikode Soc Manag Rev 1(2):55–63

Ramzi S, Afonso A, Ayadi M (2016) Assessment of efficiency in basic and secondary education in Tunisia: a regional analysis. Int J Educ Dev 51:62–76. https://doi.org/10.1016/j.ijedudev.2016.08.003

Rural Culture Committee of China Cultural Management Association (2024) Global rural crisis: rural revitalization is the solution. https://Mp.Weixin.Qq.Com/s?__biz=MzI3MDAwODMxMw==&mid=2678771354&idx=3&sn=408ee5eda80092b0f7e2889308f49317&chksm=f089c40cc7fe4d1aadaf2373a2f86372bd1d7654f52baf47e87cb169fd3f480957e02e106ea9&scene=27

Sab R, Smith SC (2001) Human capital convergence: international evidence. IMF Working Paper No.01/32

Schultz W (1987) Transforming traditional agriculture (Xiaomin Liang, Ed.). Commercial Press

Shi SB (2007) On the equalizational provision of urban-rural primary education. Shangdong University

Shukla P (2020) The role of corporate social responsibility (CSR) in development of education in India. Techno Learn 10(1&2):39–48. https://doi.org/10.30954/2231-4105.02.2021.6

Singh K, Parashar P (2024) Assessment of primary education system in rural India: a case study on Chikana Village of Tonk district in Rajasthan, India. Asian Res J Arts Soc Sci 22(1):52–62. https://doi.org/10.9734/arjass/2024/v22i1507

Stamatakis D, Petrakis PE (2006) The human capital convergence fallacy: a cross country empirical investigation. Educ Res Rev 1(3):98

Sun DP, Yu BT, Hou YX (2023) Has non-agricultural employment narrowed the gap between urban and rural education? from the perspective of human capital investment. Econ Probl 02:102–109

Tang JJ, Xu Y, Qiu HG (2022) Integration of migrants in poverty alleviation resettlement to urban China. Cities 120:103501. https://doi.org/10.1016/j.cities.2021.103501

Tang SL, Zhao FF, Zhao FF (2023) From mechanical theory to ecological theory: the philosophical turn and practical path of the integration of urban and rural compulsory education. Mod Educ Manag 10:12–23

Tu DB, Dai HQ, Deng YP (2005) An empirical study on the difference of educational development in China. Educ Sci 21(2):8-11+59

UNESDOC (2015) Education 2030: incheon declaration and framework for action: for the implementation of sustainable development goal 4. UNESDOC

Wang SF (2018) Research on the fairness index and an empirical study of the basic education resources allocation: the case of primary schools in Xicheng District, Beijing. Comp Econ Soc Syst 1:122–129

Wen J, Gu CD (2017) The rural–urban differences in resources allocation of basic education and its social consequence: base on the analysis of China's education statistics. J East China Norm Univ (Educ Sci) 35(02):33-42+117. https://doi.org/10.16382/j.cnki.1000-5560.2017.02.004

Wu YX, Wang T (2022) Dilemma and breakthrough of urban and rural education integration from the perspective of symbiosis theory. China Educ Technol 11:61–67

Xie QY (2014) Research on the balanced development of urban and rural basic education. Hunan Normal University

Xie TW, Yu WT (2021) Can education development equalize inter-provincial population urbanization? An empirical analysis based on a conditional convergence model with inter-provincial panel data. Macroeconomics 08:107–114

Xu ZG, Liu CX (2023) An optimized fractional nonlinear grey system model and its application in the prediction of the development scale of junior secondary schools in China. Sustainability 15(4):3669. https://doi.org/10.3390/su15043669

Yan XC (2020) The niche dilemma and optimization strategy of the integrated development of urban and rural Basic education. Theory Mon 11:132–139. https://doi.org/10.14180/j.cnki.1004-0544.2020.11.015

Yang XW (2019) Towards urban rural integration: value orientation and path of rural education modernization. Contemp Teach Educ 12(2):7–14

Yu X, Wang XJ (2022) Regional differences and spatial–temporal convergence of equalization of basic public education services. Stat Decis 38(17):55–60. https://doi.org/10.13546/j.cnki.tjyjc.2022.17.011

Yuan CH (2011) The study on supply mechanism of the equalization of urban–rural basal education. Ocean University of China

Yuan ZJ, Guo JWG, Yang YY et al. (2020) Analysis of the spatial structure, mechanism and social–political effects of the high-quality basic educational resources in China. Dili Xuebao/Acta Geogr Sin 75(02):318–331. https://doi.org/10.11821/dlxb202002008

Zhang D, Li X, Xue J (2015) Education inequality between rural and urban areas of the People's Republic of China, migrants' children education, and some implications. Asian Dev Rev 32(1):196–224

Zhang H (2017) Opportunity or new poverty trap: rural–urban education disparity and internal migration in China. China Econ Rev 44:112–124

Zhang W (2012) Integration of rural and urban education: epoch request of educational equity. Educ Res 33(8):13–18

Zhang ZP, Jia ZW (2016) Measuring and analysis on inclusive financial development level of Inner Mongolia: an empirical study based on variation coefficient method. J Financ Econ Theory 2:30–37. https://doi.org/10.13894/j.cnki.jfet.2016.02.005

Zhao L, Wu DT, Wang ZH et al. (2018) Spatio-temporal pattern and influencing factors of the allocation of rural basic education resources in China. Econ Geogr 38(11):39–49. https://doi.org/10.15957/j.cnki.jjdl.2018.11.005

Zheng L, Li Q, Zhang H (2023) The integration of urban and rural compulsory education in China: evolutionary trends, driving mechanisms and policy options. J Educ Stud 19(3):86–99

Zheng ZP, Yue S (2017) Does the difference of China's regional education resource allocation reduce? Based on the analysis of provincial panel data model. Res Educ Dev 37(9):28–36

Zhou YX, Song XG, Zhang LX (2019) Educational finance investment, spatial convergence and basic education equity. J Beijing Norm Univ (Soc Sci) 5:135–147

Zhou YX (2024) Changes of the school-aged population: features, trends and impacts. J Beijing Inst Educ 38(1):16–25. https://doi.org/10.16398/j.cnki.jbjieissn1008-228x.2024.01.003

Zuo H, Li S, Ge Z et al. (2023) The impact of education on relative poverty and its intergenerational transmission: causal identification based on the compulsory education law. China Econ Rev 82:102071. https://doi.org/10.1016/j.chieco.2023.102071

Author contributions

Yan Zhang (first author) contributed to conceptualization, formal analysis, draft writing and editing, project administration, and funding acquisition. Wei Yang (co-first author) contributed equally to this work, the specific contributions include methodology, validation, writing review and editing. Bo Li, Yifan Yang, and Liu Chen (co-author) contributed to the investigation. Lu Feng (corresponding author) correspondence to data curation, project administration funding acquisition, and formal analysis.

Competing interests

The authors declare no known competing interests or personal relationships that could have appeared to influence the work reported in this paper. All authors have read and agreed to the published version of the manuscript. Meanwhile, Wei Yang was a member of the Editorial Board of this journal at the time of acceptance for publication. The manuscript was assessed in line with the journal's standard editorial processes, including its policy on competing interests.

Ethical approval

The research utilized existing, second-hand data acquired from various organizations, such as Stats China and the Ministry of Education, China, among others; no direct data collection from human participants was conducted. Furthermore, our team had no direct contact with any participants, and we did not handle any personally identifying or sensitive information. Therefore, the study did not involve direct interaction with human participants, and there is no need to obtain additional ethical approval or informed consent from participants.

Informed consent

The respondents were provided information to give consents for their responses to be used in research and aggregated data to be published in research papers.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-04752-9.

Correspondence and requests for materials should be addressed to Lu Feng.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.