The Journal of
Chinese Sociology

RESEARCH   **Open Access**

Does preschool attendance affect the urban-rural cognition gap among middle school students? Evidence from China Education Panel Survey

Lei Zheng[1*], Qiuyi Weng[2] and Xin Gong[3]

* Correspondence: zhenglei@bnu.
edu.cn
[1]Faculty of Education, Beijing
Normal University, No. 19, Xinjiekou
Wai Street, Haidian District, Beijing
100875, China
Full list of author information is
available at the end of the article

Abstract

Cognitive abilities have a great impact on individuals' socioeconomic status. Among the factors that affect cognitive ability, early childhood development and preschool education are vital. However, evidence on the effect of preschool attendance on the urban-rural cognition gap from large-scale samples is almost nonexistent. Based on a nationally representative dataset from the China Education Panel Survey, this study fills this gap by presenting evidence of a significant cognition gap between urban and rural middle school students. We find a consistent positive relationship between preschool attendance and cognitive ability in grade 7 and grade 9 based on ordinary least squares and propensity score matching estimations. Results based on a Blinder-Oaxaca decomposition suggest that the differences between urban and rural students in preschool education account for 28% to 44% of the between-group cognition gap. We also simulate the effects of policy interventions meant to develop preschool education on the narrowing of the urban-rural cognition gap.

Keywords: Education equality, Urban-rural gap, Preschool education, Early childhood development, Cognition

Introduction

Cognitive ability is a developmental result of schooling and significantly impacts socio-economic outcomes and individual behaviors throughout the life cycle (Huang and Xie, 2013). In adolescence, cognitive ability can affect individuals' educational access and academic performance (Glewwe et al. 2017; Huang and Xie, 2013) and relate to intimate relationships and deviant behavior (Huang and Xie, 2013). In adulthood, cognitive ability influences individuals' career choices, work experiences, and income (Heckman et al. 2006). Cognitive ability is also a critical factor that influences intergenerational mobility, explaining 20% of intergenerational income persistence (Blander et al. 2007).

A survey of some rural and underdeveloped areas in China, however, has found that the cognitive development of rural children lags behind that of urban children (Li et al.

2017), which is an important explanation of the high dropout rate in rural middle schools and the lack of access to high school and higher education for rural students (Khor et al. 2016; Li et al. 2015; Li et al. 2017). In the long term, this gap in cognitive development will affect the accumulation of human capital, economic development, and social equity in China (Khor et al. 2016). Therefore, to promote the social mobility of rural youth and block the intergenerational transmission of poverty in rural families by improving the accumulation of human capital, we need to understand the cognition gap between urban and rural students in China, its underlying causes, and possible policy intervention tools.

Among the factors that affect cognitive ability in childhood and adolescence, early childhood development and preschool education are vital (Almond and Currie 2011a). Such experiences are early life instruments that improve individual ability, promote social mobility, and are both equitable and efficient (Heckman 2013:38). However, there is also a disparity in preschool attendance between urban and rural children in China. In 2008, the proportion of 0 to 6-year-old children in cities, towns, and villages was 18%, 21%, and 61%, respectively. However, the proportion of kindergarten children in cities, towns, and villages was 25%, 32%, and 43% (World Bank 2011). Furthermore, the benefit of preschool education to rural children is relatively low (Peng et al. 2011) due to the poor quality of care and education provided by rural kindergartens (Luo et al. 2009), the relatively low proportion of full-time teachers, and the significantly higher ratio of children to full-time teachers/staff than in urban kindergartens (World Bank 2011).

What is the impact of a preschool education experience on the cognitive levels of and the gap between urban and rural children? We lack evidence based on national samples to answer this question. Using the baseline data from the China Education Panel Survey (CEPS), this article explores the cognitive ability gap between urban and rural middle school students as it relates to preschool education. This article attempts to answer the following four questions. (1) Is the cognitive ability gap between urban and rural middle school students a nationwide phenomenon? (2) Does attending preschool in the past affect the current cognitive ability of middle school students? (3) To what extent could the cognition gap between urban and rural students be attributed to their different preschool education experiences? (4) If we implemented a nationwide policy to develop preschool education, to what extent could we narrow the cognition gap between urban and rural students?

Studying the cognition gap between urban and rural students and the impact of preschool education on that gap has important theoretical significance. Due to data limitations, few studies on social stratification and educational inequality in China have focused on cognitive ability (for example, Huang et al. 2015; Xu and Xie 2015; Zhang et al. 2015; Huang and Xie, 2013). A large number of studies on educational inequality, especially the disparity between urban and rural education, have mainly focused on the differences in school conditions, funding, and teachers at the school level, or the differences in students' educational attainment at the individual level. There are very few studies tracing the progression of educational stratification back to preschool education before compulsory school enrollment. To the best of our knowledge, very few studies explore the influence of preschool education on the cognition gap between urban and rural students by using national samples. Therefore, this article could help to expand

the scope of research on urban-rural educational inequality, clarify the process of educational stratification, and present Chinese evidence on the impact of preschool education.

Moreover, this study also has policy implications. In light of the phenomenon that the opportunity gap in higher education between urban and rural students has not narrowed and may have even expanded after the expansion of higher education, some studies have traced the roots of this gap to the inequality of opportunities between urban and rural areas at the time of the high school entrance examination (Li et al. 2015; Li 2014; Wu 2013) and the lack of cognitive preparation among rural students during preschool and primary school (Li et al. 2017). Liu (2014) also points out that the influence of cognitive ability on stratification in higher education is greater than that of family background. Therefore, exploring the causes of the urban-rural cognition gap related to preschool education could help formulate better policies to promote education equity.

Literature review

Cognitive ability and neo-human capital theory

Theodore W. Schultz and Gary S. Becker, Nobel laureates in economics, proposed the concept of human capital. This concept refers to the technology and knowledge that can be concentrated within individuals through school education, on-the-job training, health care, immigration, and other forms of investment and that can provide sustained returns (Becker 2007: XVII).

There are some limitations to the classical theory of human capital. First, the measurement of human capital is not perfectly accurate. Classical theory measures human capital as a form of investment, such as years of schooling and the length of on-the-job training (Li and Zeng 2012), making it difficult to examine differences in the quality of the final results after individuals have received the same amount of education. Some studies have found that the number of years of schooling required for the working-age population to reach the World Bank's 2nd level of basic literacy varies from 9 years to 16 years across countries (Liu 2018).

Second, the classical theory of human capital ignores the timing of capacity formation and posits that human capital investments in different stages can replace each other (Li and Zeng 2012). However, studies in cognitive neuroscience and developmental psychology have shown that there are critical and sensitive periods to develop competence (Heckman 2007).

Therefore, more and more economists have proposed to establish a "new human capital based on ability" (Hanushek 2010). Its core is composed of "skill" and "capability," including cognitive abilities and non-cognitive abilities. It attempts to explain the process by which capability is formed (Li and Zeng 2012) and emphasizes that the formation of capability should be viewed from a dynamic and life-cycle perspective (Heckman 2007).

The influence of early childhood care and education on cognitive ability
Theoretical perspectives

The early stage of life is a critical and sensitive period for ability development. At this stage, interventions for the development of infants and children and the provision of

education and care have immediate and short-term effects on their cognitive development. However, does this influence continue into adolescence and even adulthood?[1] To answer this question, various disciplines have put forward different theories, and their common ground is to dynamically view human development from a life-cycle perspective.

The fetal origins hypothesis argues that the nutritional status of the fetus during uterine development is closely related to various diseases in adulthood (Barker 1990). Since then, many studies in sociology and economics have expanded this hypothesis to examine the effects of embryonic and early environments on individual socioeconomic outcomes (Almond and Currie 2011B).

Cognitive neuroscience and developmental psychology studies on the process of neural development show that early life experiences have a lasting impact on the formation of ability. "The ages of 2 to 3 are a critical period for the development of individual oral language; the ages of 4 to 6 are the best period for children's visual identification and perception of shapes in images; the ages of 5 to 5-and-a-half are the best time to master the concept of numbers; the ages of 5 to 6 are the period when children's ability to master vocabulary develops fastest" (Pang et al. 2003). Benjamin Bloom, an American psychologist, points out that the first five years after birth are an important period of cognitive development. Approximately 50% of the intelligence measured at the age of 17 develops from the embryonic stage to the age of 4; approximately 30% develops from the ages of 4 to 8, and approximately 20% develops from the ages of 8 to 17 (Zhen et al. 2016). The "window of time" for cognitive development is relatively small. Once we understand this window of time, we can benefit significantly. Therefore, preschool education could have a continuous impact on cognitive ability.

Studies in the social sciences have also explained why early interventions and preschool education have long-term effects. Heckman (2006) summarized four key points from the research on early development interventions: (1) the formation of the brain structure and ability is affected by genetics and the environment, (2) the formation of later abilities depends on the accumulation of previous abilities, (3) cognitive abilities and non-cognitive abilities rely on each other, and (4) there is a critical period in the development of ability during which plasticity and sensitivity to environmental impacts are the strongest. This critical and sensitive period in the development of cognitive ability occurs in early childhood. Heckman and his collaborators propose the "technology of skill formation," emphasizing that the development of capabilities can be altered through conscious investments or interventions. This theory posits that the formation and development of human capability has two important features: (1) self-productivity: capability is self-reinforcing, and the abilities formed in former stages can promote the development of ability in later stages, and (2) dynamic complementarity: the productive efficiency of an intervention or investment in ability at a certain stage depends on the individual's previously acquired ability. The more ability is accumulated in the previous stage, the higher the rate of return to the intervention in the current stage is. These two features of ability production combine to form a long-term mechanism of influence: "skills beget skills" (Heckman 2007; Cunha and Heckman 2007). This theory has

---

[1]The relevant literature regards the impact of a preschool education on infants and children under 10 years old as a "short-term effect," and the impact on children over 10 years old as a "long-term effect" (Gong et al. 2016).

also been confirmed by empirical evidence that shows that early development interventions are the form of human capital investment with the highest rate of return and that their value exceeds that of subsequent investments in school education and on-the-job training (Heckman 2006).

Empirical evidence

Some review studies on early intervention programs in the USA over the past 50 years have found that by the end of such programs, the interventions have improved the cognitive abilities of the participants by approximately 0.23 standard deviations. The short-term effects of early interventions and preschool education on cognitive ability have been confirmed (Almond and Currie 2011a; Blau and Currie 2006).

However, findings on the long-term effects of preschool education are mixed. Some studies have found that preschool education has long-term positive effects on cognitive ability and educational attainment (Almond and Currie 2011a; Blau and Currie 2006). Campbell and Ramey (1995) followed participants in the Carolina Abecedarian Project in the USA and found that the IQ scores of the experimental group at the ages of 8 and 15 were 0.33 standard deviations higher than those of the control group. Heckman et al. (2013) found that the Perry Preschool Project had a long-term impact on the academic performance of the involved children. Nevertheless, some other studies have found that as time goes by, the impact of early/preschool interventions fades away (Magnuson and Duncan 2016). One explanation for this paradox is that the quality of the intervention program is of great importance. The evidence supporting the long-term effects of preschool education is mainly derived from longitudinal studies of high-quality interventions, but those studies that have not found long-term effects or have found limited long-term effects are mainly based on national/state-wide survey samples. In these samples, the quality of the preschool education programs the children attend varies. Therefore, in the long run, the effects of high- and low-quality intervention programs offset each other.

The number of relevant studies in China is relatively small, and most such studies are small sample studies. Luo et al. (2012) investigates 505 children from rural families in three provinces and six counties and finds that the level of cognitive development among children who received a formal preschool education is significantly higher than that of children without a preschool education. A study of 370 children from two townships in Guizhou Province found that rural children who attended formal kindergartens or preschool classes perform significantly better in terms of school readiness, literacy, and mathematics than those in the control group (Rao et al. 2012). Chen and Liu (2017) study a sample of 5177 15-year-old students in Shanghai and find that preschool education has a significant positive effect on mathematics, reading, and scientific literacy. Wang et al. (2017) use CEPS data to demonstrate that preschool attendance has a long-term impact on students' cognitive abilities and academic performance. However, Gong et al. (2016) employ data from the China Family Panel Studies (CFPS) and find no robust, significant relationship between preschool education and cognitive ability.

Cognitive development of urban and rural children

At present, there are few studies focusing on the cognition gap between urban and rural children in China. Based on CFPS data, several studies have found that the cognitive ability

of rural children and adolescents is significantly lower than that of their urban counterparts (Huang et al. 2015; Zhang et al. 2015; Huang and Xie, 2013). Another national survey of more than 12,000 students in grades 4 to 6 found that the cognitive ability of urban students is significantly higher than that of rural students (Tao et al. 2015).[2] The disparity in ability between urban and rural children has appeared at even earlier stages of childhood. According to a survey of families in six counties in three central and western Chinese provinces, the score on the preschool enlightenment education test for 4- to 5-year-old rural children was much lower than that of urban children of the same age (Luo et al. 2012).

The literature as mentioned above shows that cognitive ability has important social and economic value, and preschool education can advance cognitive ability. Although the long-term effects of preschool education are uncertain, there are at least short-term effects. Therefore, this study proposes several hypotheses to be verified: (1) there is a significant gap in cognition and significant inequality in preschool attendance between urban and rural students, (2) preschool education has long-term effects on the cognitive ability of middle school students, (3) the differences in preschool experiences between urban and rural students is an important explanator of the current cognition gap, and (4) policies that popularize preschool education in rural areas can narrow the urban-rural cognition gap to an extent.

Methodology

Data

This study uses the baseline data from the China Education Panel Survey from 2013 to 2014. This survey targeted middle-school students in grade 7 and grade 9. Through a multi-stage stratified sampling method, four sampling units—including county (district), school, class, and student/parent/headteacher/main subject teacher/school leader—were sampled in turn. Data on 19,487 middle school students from 112 schools and 438 classes in 28 counties (districts) were collected from the baseline survey.[3]

Variables

Definition of urban and rural observations

In this study, the "urban" and "rural" samples are defined according to the household registration (*hukou*) of the students in the survey.[4] The measurement of the control variable "household registration" defined in a later section is the same. This definition is mainly based on the following two considerations. First, in China, access to many basic public services, including preschool education, is linked to household registration. Defining the urban and rural samples according to household registration rather than the location of the school ensures that the household registration status of the students in the survey was the same status they had when they were facing the different options

---

[2]During the revision to respond to the peer reviewers, we identified two more relevant studies. Zhao et al. (2017) and Jiang (2017) both used CEPS data to demonstrate that there was a significant gap in cognitive ability between urban and rural students, but the two studies did not examine the contribution of a preschool education to the cognitive gap.

[3]For details, please refer to the User Manual for the Baseline Data of the China Education Panel Survey (CEPS) (http://www.cnsda.org/index.php?r=projects/view&id=72810330).

[4]On the basis of the variables for the students' household registration provided by CEPS, "non-agricultural household registration" and "resident household registration" are combined into "non-agricultural household registration," which is used to define "urban students" in this study. "Rural students" are defined as holders of an "agricultural household registration."

for preschool education (whether to participate in preschool education or not). This is certainly true for students whose household registration status remained unchanged. For those rural students who had changed from agricultural household registration to urban household registration after they had obtained their preschool education, their household registration statuses at these two-time points would be different. However, the size of this group of students is relatively small (n = 460), and its impact is not significant.

Second, many students have agricultural household registration (migrant children with a nonlocal and agricultural household registration) studying in urban schools. If the samples were defined by the school's location, we would obtain an urban group composed of students with two different household registrations. Although these migrant children are currently studying in urban schools, many of their experiences before migrating to the city were quite different from those of students with urban household registrations. Hence, the factors influencing their cognitive abilities would be quite different. Although the definition of rural and urban samples may result in an urban group with some "rural" students who have converted to non-agricultural household registration, the size of this subsample is far smaller than the number of migrant students who have gone to the city to study but have not changed their household registration. Compared with a definition for urban and rural samples according to school location, a "purer" urban-rural grouping can be obtained by using household registration.

Dependent variable
CEPS designed a set of cognitive test questions for grade 7 and grade 9. This study uses the standardized score after conversion as the dependent variable to measure students' cognitive ability.

Independent variables
CEPS asked respondents whether they had attended kindergarten or preschool after the age of three. We construct the core independent variable "received preschool education (yes = 1)" based on this question.

The length of preschool education (the duration of the stay in kindergarten) may also be different between urban and rural students. Therefore, when we compare the preschool education experiences of urban and rural students, we also investigate the length of the kindergarten experience. According to national policy, kindergartens accept children three to six years old. CEPS asked the students whether they had attended a kindergarten or preschool class after the age of three, and it asked them how old they were when they had started kindergarten or preschool. A small number of students reported that they were admitted into kindergarten before they were 3 years old or after they were 6 years old. To ensure the consistency of the answers to these two questions and their conformity with national policy, we remove the observations whose admission age was younger than 3 years old or older than 6 years old. Then, we assume that once children enter kindergarten, they continue to attend kindergarten until they enter primary school. The duration of their stay in kindergarten is calculated by subtracting the

age of kindergarten admission from the age at primary school admission reported by the students.

Other control variables

Following existing studies (Gong et al. 2016; Liu and Xie, 2015; Wang et al., 2017), we control for other factors that may affect students' cognitive abilities. These factors include:

(1) Individual-level variables: gender (male = 1), age, ethnicity (Han = 1), household registration (agricultural household registration = 1), physical disability (various disability statuses = 1), whether a student is a single child (single child = 1).

(2) Family-level variables: the years of education of the most-educated parent, the subjective self-evaluation of the family's current economic status (dummy variables for "family's economic status is middle class" and "family's economic status is wealthy," with "family's economic status is poor" as the reference group), size of the family book collection (dummy variables for "collection of books is medium-sized" and "collection of books is relatively large or large," with "collection of books is small" as the reference group), whether parents are at home (dummy variables for "one parent is not at home" and "both parents are not at home," with "both parents are at home" as the reference group).

(3) School-level variables: the local ranking of schools (dummy variables for "top schools" and "above-average schools," with "below-average schools" as the reference group), and the school type (private school = 1).

(4) Regional-level variables: since the level of regional development affects access to educational opportunities and the educational performance of individuals, we also control for the average length of schooling of the residents in the district/county and for the region where the school is located (dummy variables for "central region" and "western region," with "eastern region" as the reference group).

We limit the analytical sample to respondents with no missing values for any variable. The sample for the analysis ultimately includes 17,748 students.[5] Given that grade 7 and grade 9 use different cognitive tests and their total scores are different, it is reasonable to analyze these two grades separately.

The descriptive statistics of the variables[6] show that there are significant differences in cognitive abilities and preschool education experiences (admission rate and duration of the kindergarten experience) between urban and rural students in grade 7 and grade 9 (see Tables 1 and 2 for details). Compared with urban students, a higher proportion of rural students have disabilities; a higher proportion of rural students have siblings; the education level of rural students' parents is lower; their family economic status is worse; the proportion of rural students whose family book collection is medium-sized is generally higher, while the proportion whose family book collection is relatively large or large is significantly lower; a higher proportion of rural students have parents who are not at home (especially when both parents are not at home); a higher proportion of

---

[5]The sample size is 15,191 when analyzing the difference in the duration of the kindergarten experience between urban and rural students.
[6]To save space, only the variables with statistically significant differences are briefly explained here. Detailed results can be obtained from the authors.

Zheng *et al. The Journal of Chinese Sociology* (2021) 8:14

Page 9 of 22

Table 1 Cognitive ability of urban and rural students

Grade	Sample		N	Mean	Diff.	Effect size
7	Nationwide	Urban	4332	0.154	0.248***	0.290
		Rural	5104	− 0.094		
	West region	Urban	1181	0.120	0.327***	0.399
		Rural	1111	− 0.207		
9	Nationwide	Urban	3637	0.215	0.362***	0.438
		Rural	4675	− 0.147		
	West region	Urban	1046	0.243	0.478***	0.570
		Rural	973	− 0.236		

Notes: (1) $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

rural students attend above-average schools, while a lower proportion attends top schools; a higher proportion of rural students attend private schools; a higher proportion are from the central region, but a lower proportion is from the western region. Compared with their urban counterparts, the rural students live in the district/county with a lower average education level of local residents.

Empirical strategies

Descriptive statistics used to describe the differences between urban and rural areas

We use descriptive statistics to analyze whether there is a significant gap between urban and rural students' cognitive levels and preschool education experiences and whether this gap is a nationwide phenomenon.

Table 2 Preschool experience of urban and rural students

Grade	Sample		N	Mean	Diff.	Effect size
1. The proportion of preschool attendance						
7	Nationwide	Urban	4332	85.66%	8.22%***	0.212
		Rural	5104	77.45%		
	West region	Urban	1181	87.47%	13.39%***	0.346
		Rural	1111	74.08%		
9	Nationwide	Urban	3637	84.11%	9.71%***	0.239
		Rural	4675	74.40%		
	West region	Urban	1046	85.28%	14.36%***	0.354
		Rural	973	70.91%		
2. The duration of preschool education (the years of staying at kindergarten)						
7	Nationwide	Urban	3688	2.42	0.45***	0.328
		Rural	4317	1.97		
	West region	Urban	1011	2.47	0.75***	0.577
		Rural	954	1.71		
9	Nationwide	Urban	3145	2.31	0.57***	0.422
		Rural	4041	1.74		
	West region	Urban	919	2.34	0.91***	0.722
		Rural	857	1.43		

Notes: (1) $^*p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$.

Ordinary least squares (OLS) model and propensity score matching (PSM) estimation to identify the effect of preschool education

To strictly identify the long-term effect of preschool attendance on cognitive ability, we would need to use a randomized experiment and continuous tracking of the students. However, such longitudinal data is almost unavailable in China. Therefore, most studies based on survey data have first conducted OLS estimations and then adopted PSM (e.g., Gong et al. 2016; Chen and Liu 2017) or instrumental variable (e.g., Borraz and Cid 2013) estimations to reduce estimation bias. Since the data in this paper is cross-sectional and appropriate instrumental variables are unavailable, we use an OLS model and PSM to estimate the effect of preschool education.

The OLS model is shown in formula (1):

$$Cognitive_i = \alpha + \beta \cdot Preschool_i + \gamma \cdot X_i + \varepsilon_i \qquad (1)$$

$Cognitive_i$ is the standardized cognitive test score of student i. $Preschool_i$ is a dummy variable indicating whether the student received a preschool education, and X_i is a set of control variables. Since the CEPS sampling process includes clustering and the homogeneity in students from the same school is high, the standard errors are adjusted by clustering at the school level. OLS estimation is conducted by grade. In each grade, we first estimate the model on three nationwide samples—the urban sample, the rural sample, and the pooled sample; then, we focus on the western region and estimate the model on the above three samples but restricted to those from the western region. In sum, 12 models are estimated.

Preschool attendance is self-selected. There are significant differences in many characteristics between the preschool attendees (treatment group) and the preschool non-attendees (control group). For example, the treatment group is younger; it has a higher proportion of single child; it has a higher proportion of students of Han ethnicity; and it has a lower proportion of students with a disability. The family socioeconomic status of students in the treatment group is higher (parents' education levels are higher, family economic status is higher, the family collection of books is larger, and the proportion of absent parents is lower). The quality of the schools the treatment group students attend is higher, and the level of education attainment in the region from which treatment group students come is higher.

Although the OLS model controls for these confounding variables, it is difficult to ensure that the students in the treatment and control groups are comparable in terms of the different combinations of each confounding variable as the number of confounding variables increases. PSM can solve this problem. PSM can simulate one (or more) control case that is close to the treatment case in all respects and thus obtain comparable cases by estimating a one-dimensional propensity score based on multiple confounding variables and controlling for it (Hu 2015:36–37); this method does not depend on the specification of the regression model. Therefore, it is often employed to conduct causal inference with non-random survey data as a robustness test for OLS estimation (Gong et al. 2016). In this study, PSM is conducted as follows. First, the probability of individual receiving preschool education, that is propensity score, is estimated by using a probit model. When estimating the propensity score, we control for a set of individual, family, and regional level covariates that may affect students' preschool attendance.

Second, the cases whose propensity scores are in the common support are matched. Four matching methods are adopted to ensure the robustness of the estimation results: caliper matching (caliper = 0.01), nearest-neighbor matching within a caliper (caliper = $0.25\,\hat{\sigma}_{pscore}$, k=4), kernel matching (using the default kernel function and bandwidth), and local linear regression matching (using the default kernel function and with band-width = 0.5).

Third, by determining whether the standardized differences in the variables between the two groups are less than 10% after matching, we can validate whether the matching has effectively balanced the data.

Finally, the average treatment effect on the treated (ATT) is calculated by using the matched samples; that is, for those who have received a preschool education, the ATT is the average difference in their cognitive ability between two states: receiving a preschool education (factual) and not receiving a preschool education (counterfactual).

In the PSM estimation process, 12 different samples are estimated, as previously described for the OLS model.

The contribution of preschool education to the urban-rural cognition gap based on a Blinder-Oaxaca decomposition

Based on the OLS model, the Blinder-Oaxaca decomposition is used to explain the mean difference between groups. The mean difference between groups can be attributed to two effects: the difference in the "endowments" between the two groups; the difference in the "rate of return" to those endowments. The term "endowment" refers to the value of each independent variable that influences the dependent variable, while the term "rate of return" refers to the effect of each independent variable. For example, in this study, the influence of preschool attendance on the urban-rural cognition gap has two aspects: (1) access to kindergarten is different between the two groups and (2) the influence of preschool attendance on cognitive ability (the rate of return to preschool) is different between the two groups.

We conduct OLS estimations similar to formula (1) for the urban and rural samples separately. For the sake of simplicity, we do not list the variable for preschool education (*Preschool*) separately but put it into the covariate X. Then, the mean difference in cognitive ability between the urban and rural groups can be expressed by formula (2):

$$
\begin{aligned}
\overline{Cognitive_u} - \overline{Cognitive_r} &= \left(\widehat{\alpha_u} + \widehat{\beta_u} \cdot \overline{X_u}\right) - \left(\widehat{\alpha_r} + \widehat{\beta_r} \cdot \overline{X_r}\right) \\
&= (\overline{X_u} - \overline{X_r})\widehat{\beta_r} + \overline{X_u}\left(\widehat{\beta_u} - \widehat{\beta_r}\right) + (\widehat{\alpha_u} - \widehat{\alpha_r})
\end{aligned} \tag{2}
$$

The subscripts *u* and *r* denote the urban and rural samples, respectively. The first item on the right side of formula (2) is the "explained part" or "endowment effect," which measures the contribution of the differences in endowments to the urban-rural cognition gap. The second and third items are the "unexplained parts" or "coefficient effects," which measure the contribution of the differences in the regression coefficients and intercept to the urban-rural cognition gap.

Policy simulation

Based on the effect of preschool education and the difference in urban and rural preschool education opportunities estimated with the OLS model, we simulate the effect

of a policy to popularize preschool education on the narrowing of the cognition gap between urban and rural students. See the results section for details.

Results and discussion

Differences in cognitive ability and preschool experiences between urban and rural students

The urban-rural cognition gap

The urban-rural cognition gap has the following two features (see Table 1 for details). First, the differences in cognitive ability between urban and rural middle-school students are universal and nationwide. The disparity in cognitive ability between urban and rural students in grade 7 is 0.248 standard deviations, and that in grade 9 is 0.362 standard deviations. This difference is not only significant ($P < 0.01$) but is also a medium-sized effect (Cohen's $d > 0.2$). This may be because the grade 7 students were admitted into the school later than the students in grade 9. With the development of the education system, the gap between urban and rural areas has been narrowed.[7]

Second, the urban-rural cognition gap is a typical western-region problem. The results show that the disparity in cognition between urban and rural students is most prominent in the western region when the nationwide sample is divided into eastern, central, and western subsamples for analysis.[8] The difference in cognitive ability between urban and rural students in grade 9 in western China is as wide as 0.478 standard deviations, which is a very large difference (Cohen's $d > 0.5$). It may be related to the population distribution and resource concentration in the region. On the one hand, the western region in China is vast and sparsely populated, and resources are heavily concentrated in urban areas. On the other hand, rural areas in the western region are extremely underdeveloped, and most nationally designated poor counties are located in this region. The striking urban-rural gap in the western region is the consequence of the nexus between resource agglomeration in urban areas and underdevelopment in rural areas. This implies that to narrow the disparity between urban and rural areas, policies should be focused on the western region. Therefore, we examine the national sample and the western region subsample separately in the following sections.

Preschool experiences of urban and rural students

There is a significant difference in preschool experiences between urban and rural students. First, the proportion of urban children who have received a preschool education is significantly higher than that of rural children, and the gap is medium-sized (Cohen's $d > 0.2$). Similarly, this gap is more pronounced among students in grade 9 and those in the western region (see Table 2, Part 1).[9]

---

[7]This result cannot be interpreted as indicating that the cognitive gap among grade 7 students will widen as they progress into grade 9 because in the cross-sectional data that are used here, the grade 9 and grade 7 students are not in the same cohort. Further tests on this issue would be possible if longitudinal data were available. We appreciate the reviewer offering this suggestion.

[8]To save space, we do not report the results for the eastern and central subsamples.

[9]In the CEPS data, no matter how many years of preschool education children have received and no matter how old they are when they enter kindergarten, they are counted as having received a preschool education. Therefore, the proportion of children who have received a preschool education that is calculated based on this dataset is higher than the gross enrollment rate for preschool education that is released by educational administrations. To save space, the results for the eastern and central subsamples are not shown in the article.

Second, the duration of preschool education is longer among urban children. The World Bank found that approximately 48% of rural children in China had received one year of preschool education, rather than the 3 years of kindergarten generally received by urban children (World Bank 2011). Our data show that 47.58%, 35.84%, and 11.67% of urban students entered kindergarten at the ages of 3, 4, and 5, respectively, with an average enrollment age of 3.76. The proportion of rural students entering kindergarten at the ages of 3, 4, and 5 was 33.21%, 37.06%, and 20.05%, respectively, with an average enrollment age of 4.08 years. The second part of Table 2 shows that the difference in the duration of preschool education between urban and rural students is not only significant but is also remarkable in size. Especially in the western region, the gap is extremely large (Cohen's $d > 0.5$).

The effect of preschool attendance on cognitive ability
Results of the OLS model
The results of the OLS regression are as follows (see column 2 of Table 3 for details). First, there is a significant and robust relationship between preschool attendance and cognitive ability. On average, in the nationwide sample (the western sample), the cognitive ability of the students who had received a preschool education is 0.1 (0.22) standard deviations higher than that of the control group. This demonstrates that preschool education has long-term effects on cognitive ability. Compared with the coefficients of the other variables, the coefficient of the preschool education variable is not only robust but is also relatively large in magnitude. To control for inter-school heterogeneity, we also estimated a school fixed-effects model and a hierarchical linear model (HLM) and found that the results do not change significantly.[10]

Second, the effect of preschool attendance is more prominent in the western subsample. It is partly because other factors that may affect the development of cognitive ability are generally scarce in the western region. For example, the level of paternal education is lower, and family book collections are smaller. As a result, the impact of preschool education on cognition is substantial.

Third, the preschool effect is larger in magnitude for the grade 7 sample than the grade 9 sample. It may be due to the cohort effect: the two grades do not include the same observations, so there are differences in the sample distribution. It may also be due to the period effect: compared with the grade 7 students, the grade 9 students received a preschool education earlier and faced different education policies and environments. For example, the quality of the preschool education received by the grade 9 students may be lower than that of the grade 7 students. To be sure, it may also be due to the age effect: the older the age of the students or the higher the grade is, the more prominent the roles of the other factors affecting cognitive ability are, which, to a certain extent, offsets the role of preschool education. It is necessary to use longitudinal data to test the above hypothesis strictly. In short, this result cannot simply be interpreted as indicating that the effect of preschool education fades away with an increase in grade level.

Fourth, for grade 7 students, the effect of preschool education is more prominent among urban students. Especially in the western region, the estimated effect for the

---

[10]To save space, the results of these models are not shown in this article.

Table 3 The effect of preschool attendance on cognitive ability

Sample		OLS	PSM			
			Caliper	Neighbors within caliper	Kernel	Local linear regression
Grade 7	Nationwide pooled	0.162***(0.024)	0.172***(0.025)	0.183***(0.027)	0.178***(0.025)	0.180***(0.032)
	Nationwide urban	0.203***(0.042)	0.205***(0.044)	0.252***(0.046)	0.203***(0.043)	0.221***(0.055)
	Nationwide rural	0.136***(0.025)	0.135***(0.029)	0.162***(0.032)	0.142***(0.029)	0.136***(0.037)
	Western pooled	0.223***(0.044)	0.227***(0.058)	0.247***(0.063)	0.241***(0.055)	0.264***(0.077)
	Western urban	0.305***(0.068)	0.284***(0.093)	0.279***(0.096)	0.317***(0.084)	0.345***(0.112)
	Western rural	0.164***(0.047)	0.157**(0.067)	0.175**(0.071)	0.175***(0.064)	0.190**(0.079)
Grade 9	Nationwide pooled	0.098***(0.023)	0.100***(0.025)	0.085***(0.027)	0.111***(0.025)	0.130***(0.033)
	Nationwide urban	0.097**(0.039)	0.079*(0.047)	0.053(0.050)	0.087*(0.045)	0.103*(0.058)
	Nationwide rural	0.101***(0.029)	0.111***(0.029)	0.112***(0.031)	0.128***(0.028)	0.121***(0.037)
	Western pooled	0.116**(0.045)	0.101*(0.060)	0.136*(0.063)	0.117**(0.059)	0.212***(0.072)
	Western urban	0.141*(0.071)	0.096(0.111)	0.062(0.113)	0.064(0.106)	0.116(0.132)
	Western rural	0.113*(0.061)	0.139**(0.066)	0.151**(0.071)	0.166**(0.065)	0.198**(0.080)

Notes:(1) Each coefficient is from a separate model; (2) *p<0.1, **p<0.05, ***p<0.01; (3) All control variables are included in OLS models and the clustered standard errors are in parentheses; (4) The Average Treatment effects for the treated (ATT) are reported for PSM estimation and the standard errors are in parentheses

urban sample is almost double that for the rural sample. It may be due to the relatively low quality of preschool education in rural areas, limiting the rate of return to preschool education for rural children.

Fifth, among other influencing factors, the effect of parental education and family cultural capital is robust and significant. Comparing the estimated coefficients for the family economic status variables, we can see that culture is more important than money for children's cognitive development. In most models, having a disability has a significantly negative effect on cognition. The impact of school quality on cognitive ability is not robust, as it depends on the specific sample examined.

Results of the propensity score matching estimation

First, we use a probit model to estimate each individual's propensity to receive preschool education.[11] The results show that the older a student is, the less likely she or he is to have received preschool education. Additionally, the better the family economic situation is, the larger the book collection is, and the higher the average education level in the district/county is, the more likely individuals are to receive preschool education.[12] The presence of a physical disability, parental education levels, and ethnicity are also related to an individual's access to preschool education (but the effect is dependent on the specific sample).

Next, four matching methods are used to estimate the ATT. No matter which matching method is used, most observations can be matched successfully.[13] The standardized

---

[11]To save space, the prediction result of propensity score is not shown in this article.

[12]These individual or household characteristics are not predetermined before the preschool investment decision is made. However, it is reasonable to speculate that these characteristics do not change greatly over time.

[13]Most observations that did not match successfully were observations from the western rural areas in grade 7. A total of 51 of the 1111 observations were not matched successfully. For details, see Table A1-1~Table A1-4 of online supplementary materials.

differences for almost all variables after matching were less than 10%, which validates that the matching effectively balanced the data.[14]

Columns 3–6 of Table 3 report the ATT among preschool attendees, and these results do not diverge from the results of the OLS estimation. The only difference is that for the grade 9 students, the effect of preschool education among urban students is smaller than that among rural students and is no longer significant. This may partly be due to the difference between the grade 9 urban and rural samples.

The effect of preschool attendance estimated in this study is close to that estimated in other studies. For example, Magnuson and Duncan (2016) summarize various early/preschool intervention programs in the USA over the past 50 years and find that these programs improved the cognitive ability of participants by approximately 0.23 standard deviations. Chen and Liu (2017) use the PISA data for Shanghai, China, in 2012 and find that the effect of preschool education on mathematics, science, and reading literacy was 0.08 to 0.17 standard deviations.

However, Gong Xin et al. (2016) use CFPS data and find that preschool attendance and children's cognitive abilities do not have a robust significant relationship. There are two possible explanations for this. First, the two studies measure cognitive ability differently. The cognitive abilities measured in the CFPS2010 data, which Gong et al. (2016) used, are "the verbal ability and calculating ability obtained from school education, and most items are designed according to the content of primary and secondary school curricula. Therefore, the cognitive ability measured by CFPS2010 is highly related to education levels" (Huang and Xie, 2013:126). This measure of cognitive ability is similar to the "crystallized intelligence" defined by psychologist Raymond Cattell, which refers to the kind of cognitive ability acquired through experiences, such as language ability, ability to judge, and association ability, and is highly related to school education. In the CEPS data used in this paper, the content of the cognitive ability test "does not involve the specific memorization of knowledge taught by a school curriculum, but measures students' logical thinking and problem-solving ability" (The User Manual for the Baseline Data of the China Education Panel Survey:8).[15] This is similar to Cattell's definition of "fluid intelligence"—a physiology-based cognitive ability, such as perception, memory, speed of calculation, and ability to reason. Fluid intelligence develops as the nervous system matures, reaches its peak at middle age, and then decreases slowly. Therefore, the two cognitive abilities have different connotations and thus distinct effects. The former measures the cognitive ability that is mainly derived from the knowledge taught in the school curriculum, which is greatly affected by the current school. The latter measures a more "internal" cognitive ability, which is highly related to the development of the nervous system. Early education aims to intervene precisely in the development of the human cognitive nervous system during its sensitive and critical period, so it has a greater impact on fluid intelligence in this sense.

Second, the sample sizes of the two studies are quite different. The sample size in Gong's study is less than 3000, while the sample size in this study is 17,748. The sample

---

[14]To save space, the balance test results are not shown in this article. See Table A2 of online supplementary materials for details.

[15]For details, please refer to the User Manual for the Baseline Data of the China Education Panel Survey (http://www.cnsda.org/index.php?r=projects/view&id=72810330).

size of the former study is small, resulting in low levels of variation among individuals and relatively less significant estimation results.

The contribution of preschool attendance to the urban-rural cognition gap

We examine the contribution of preschool attendance to the urban-rural cognition gap with a Blinder-Oaxaca decomposition. The results are shown in Table 4.

In general, the urban-rural cognition gap is attributed to differences in endowments (preschool education is an exception). It implies that the disparity in cognitive ability between urban and rural students is mainly due to differences in the characteristics of the two groups. The results of the decomposition show that the education level of parents and the cultural capital of families are vital determinants of cognitive ability. Whether students have disabilities and family economic status contribute little to the cognition gap. School ranking only makes a considerable contribution to the gap within the grade 7 sample. To an extent, the education level of parents and the cultural capital of families reflect parents' ideas, habits, and knowledge regarding raising their children. The huge difference between urban and rural families in this respect explains the cognition gap to an extent, while the difference in household economic conditions is not important.

Preschool attendance is an important explanator of the urban-rural cognition gap in grade 7. It accounts for 27.55% of the cognition gap among grade 7 students in the nationwide sample and 44.37% in the western sample. Specifically, the contribution of preschool attendance mainly comes from the coefficient effect. In other words, although the gap in preschool opportunities between grade 7 urban and rural students explains part of the cognition gap, the more important explanation is that the rate of return to preschool education is different for urban and rural students.

However, preschool attendance is not an important contributor to the urban-rural cognition gap in grade 9: its contribution is less than 10%. In particular, the contribution of preschool attendance to the urban-rural cognition gap among grade 9 students in the nationwide sample is almost equal to zero. This is because (1) the proportion of urban students receiving a preschool education in grade 9 is higher than that of rural students (see Table 1) and the endowment effect is positive; (2) the difference in the rate of return to a preschool education between grade 9 urban and rural students is rather small, even for urban students and the coefficient effect is negative; and (3) the contribution of preschool education to the urban-rural cognition gap is the sum of the

Table 4 The contribution of preschool attendance to the cognition gap

	Grade 7		Grade 9	
	Nationwide pooled	**Western pooled**	**Nationwide pooled**	**Western pooled**
Explained part	85.07%	86.05%	108.90%	115.87%
Unexplained part	14.93%	13.95%	− 8.90%	− 15.84%
Total	100.00%	100.00%	100.00%	100.03%
Preschool attendance				
Explained part	4.52%	6.73%	2.71%	3.39%
Unexplained part	23.03%	37.64%	− 1.08%	5.02%
subtotal	27.55%	44.37%	1.63%	8.41%

above two parts, with the negative effect offsetting the positive effect and resulting in a very small contribution. The urban-rural cognition gap in grade 9 is mainly explained by the following factors: whether the student is a single child, parental education levels, and the size of the household book collection. Given that the total contribution of all factors is 100%, the increase in the contribution of these factors "dilutes" the effect of preschool education to a certain extent.

Since the grade 7 and the grade 9 students are not from the same cohort, the difference in the contribution of preschool education to the cognition gap between the two grades should be explained as a difference between the samples. It cannot be interpreted as implying that when the grade 7 students progress to the next grade, the contribution of preschool education to the cognition gap becomes as small as that in the current grade 9 sample. The differences between the two samples may come from a "cohort effect": the differences in family background and the two groups' initial cognitive ability levels. It may also come from the period effect: the education system itself has changed over time and thus has heterogeneous impacts on students. For example, the quality of the preschool education received by the two groups is different.

The influence of a policy to develop preschool education on the narrowing of the urban-rural cognition gap

The results of the decomposition demonstrate that the urban-rural cognition gap is mainly caused by differences in preschool education experiences, the education level of parents, family cultural capital, and the quality of the current school. To narrow the gap in cognitive ability between urban and rural students, we need to eliminate the differences in these characteristics between urban and rural students. Apparently, aside from preschool education experiences, these factors are difficult or even impossible to affect through public policy. Policies to develop preschool education, which aim to expand the enrollment in preschool and improve the quality of care and education in preschools, are plausible. Since preschool attendance is a prominent contributor to the urban-rural cognition gap in grade 7 but makes a relatively small contribution to the cognitive disparity in grade 9, we use the grade 7 sample to simulate the effect of a universal preschool education policy on the narrowing of the cognition gap.

Ideally, if the policy ensured that urban and rural children had the same opportunity to receive preschool education and the effect of that preschool education on them was the same, the cognition gap would be reduced by 27.55% (in the nationwide sample) or 44.37% (in the western sample). Thus, the effect of the policy intervention would be remarkable, given that the other factors remained the same.

There are differences in preschool education experiences for urban and rural children, including access to, the duration of, and the rate of return to such an experience. What ought to be the specific policy initiative chosen from among alternatives such as increasing opportunities, extending durations, and improving the benefits of preschool for rural children? Following Magnuson and Waldfogel (2005), we simulate the effects of different intervention schemes. Since increasing the rate of return is related to improving the quality of preschool education, we consider just the effects of the first two schemes, which are easier to operationalize.

Increase preschool education opportunities for rural children

If a policy can increase the proportion of rural children entering kindergarten to the same proportion as that in urban areas, it means that the proportion of rural children receiving a preschool education would increase to 85.66% (in the nationwide sample) or 87.47% (in the western sample). When other factors remain unchanged[16], a policy that increases opportunities could reduce the urban-rural cognition gap by 4.44% (in the nationwide sample) or 6.73% (in the western sample) (see Panel A of Table 5).

However, this policy effect is relatively small compared with the overall contribution of preschool education to the cognition gap revealed by the decomposition results. It shows that expanding enrollment in preschool is not sufficient to eliminate the disparity in cognition between urban and rural students. Instead, we should make an effort to ensure that rural children receive a high-quality preschool education of sufficiently long duration.

Prolonging the duration of kindergarten for rural children

Some studies have found that the duration of preschool education has a positive effect on children's cognitive abilities (Chen and Liu, 2017). We employ an OLS model (that includes the same control variables) and find a significant positive correlation between kindergarten duration and cognitive ability. The effects of duration are 0.077 and 0.062 standard deviations for the urban and rural students (in the nationwide sample), respectively, and 0.055 and 0.052 standard deviations for the urban and rural students (in the western sample), respectively.[17] Therefore, prolonging the duration of preschool education for rural children could also narrow the urban-rural cognition gap.

The simulation results in Panel B of Table 5 indicate that if the length of time rural children spend in kindergarten is extended to the same level as that of urban children, the cognition gap could be reduced by 11.29% (nationwide sample) or 12.23% (western sample), even if other factors remain unchanged.

Taking the two initiatives together, if rural children's access to kindergarten is guaranteed and their kindergarten experience is prolonged, the gap in cognitive ability between urban and rural students could be reduced by 15.73% (nationwide) or 18.96% (western regions), even if other factors are held constant.

It is worth noting that the effects of the policy intervention are underestimated. First, the respondents to the CEPS are all enrolled students, and dropouts of the same age are excluded. The average socioeconomic status of this sample of enrolled students is higher than that of the full youth population, and this is especially true for the rural population. Therefore, the real gap between urban and rural children's preschool education experiences may be larger. This implies that if urban children's preschool education conditions are used as the benchmark, there is more space to improve rural children's preschool education opportunities and prolong their kindergarten experience, and the corresponding policy effects would be stronger. Second, since the difference in preschool quality between urban and rural areas is much greater than the differences in

---

[16]Besides assuming that the distribution of each influencing factor in the model remains constant, it is necessary to assume that the current proportion of urban and rural students remains unchanged. We appreciate the reviewers for pointing this out.

[17]The p values for these coefficients are all 0.000. To save space, the specific results are not shown in this article.

Table 5 The effects of policy simulation

Panel A			
Policy 1: Increasing preschool opportunity for rural children ①	Effects of preschool attendance ②	Average improvement③(=①×②)	Gap narrowed④(=③/cognition gap)
Nationwide: 77.45%→85.66%	0.136	0.011	4.44%
Western: 74.08%→87.47%	0.164	0.022	6.73%
Panel B			
Policy 2: Extending preschool duration for rural children⑤	Effects of preschool duration⑥	Average improvement ⑦(=⑤×⑥)	Gap narrowed⑧(=⑦/cognition gap)
Nationwide: 1.97 years→2.42 years	0.062	0.028	11.29%
Western: 1.71 years→2.47 years	0.052	0.040	12.23%
Panel C			
Total effect of the two policies(=④+⑧)			
Nationwide			15.73%
Western			18.96%

Notes: (1) The data on the increase of preschool opportunity and the extension of duration are from Table 2; (2) The effects of preschool attendance are from Table 3; (3) The effects of preschool duration are estimated on the Grade 7 rural sample of nationwide and western region respectively; (4) The data on cognition gap is from Table 1

education opportunities or education duration, the effect of a policy to develop preschool education that targets equal quality would be stronger. Finally, as the development of ability is a dynamic process of continuous accumulation, the earlier an intervention is, the more significant its effect is. If an intervention policy could help rural children to receive the same level of preschool education as urban children in their infancies, the initial cognition gap between the two groups would be smaller, resulting in a correspondingly smaller gap in cognitive ability in later life cycle. Because of these three reasons, the policy effect simulated in this paper can be understood as a relatively conservative estimation and a lower bound.

Conclusions

Using the baseline data from the China Education Panel Survey, this paper finds a significant gap in the cognitive ability and preschool education experiences of urban and rural students, especially in the region of western China. The results of the OLS regression and the PSM estimation demonstrate that preschool attendance can significantly affect the cognitive ability of middle school students. The Blinder-Oaxaca decomposition reveals that preschool education is a vital explanator of the cognition gap between urban and rural students in grade 7, but it does not contribute much to the cognition gap between urban and rural students in grade 9. The policy effect simulations show that certain policies can reduce the cognition gap between students from urban and rural areas by increasing the opportunities to receive a preschool education for rural children and prolonging their kindergarten experience.

Given the findings in this study, first, a policy to develop rural preschool education should pay equal attention to guaranteeing access, prolonging the duration, and improving quality. Simply increasing rural children's access to preschool education cannot help reduce the cognition gap between urban and rural students.

Second, policies to develop preschool education should be focused on the western region. In other words, the imbalance between urban and rural areas in terms of education and even social and economic development as a whole is mainly a western China

problem. If such policies focused on the western region, especially impoverished areas and the old revolutionary base, minority ethnicity areas, border areas, and underdeveloped areas, the nationwide gap in children's cognitive development between urban and rural areas would be effectively narrowed.

Finally, the target of a policy to develop preschool education should be a "combination of medical care and education" and involve "cooperation between families and kindergartens" (Li 2011). This study finds that disabilities limit children's access to preschool education and affect their cognitive ability. Hence, a policy to develop preschool education should integrate education, care, and nutrition (Li 2011). The cultural capital of the family and the education level of parents are also found to have a strong influence on children's cognitive abilities. Many rural families do not have proper environments of parental reading, and rural parents do not have an awareness of or thoughts about early/preschool interventions for their children. Therefore, a policy to develop preschool education would not only rely on institutions to provide early development services but would also need to provide advice and guidance to parents to improve the quality of family care and education.

This study has some limitations. Although we have adopted various methods to identify the effect of preschool attendance on cognitive ability, the estimated effects in this study are not strict causal effects. There is still limited research on the effect of preschool education in China, and more follow-up studies are needed to further understand different aspects of this issue, such as the size of the effect of preschool education, whether the effect changes with age, and what the mediating mechanism of the effect is.

Abbreviations
CEPS: China Education Panel Survey; CFPS: China Family Panel Studies; OLS: Ordinary least squares; PSM: Propensity score matching; ATT: Average treatment effect on the treated

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s40711-021-00150-1.

Additional file 1: Table A1-1. Sample sizes after matching (caliper matching). **Table A1-2.** Sample sizes after matching (nearest-neighbor matching). **Table A1-3.** Sample sizes after matching (kernel matching). **Table A1-4.** Sample sizes after matching (local linear regression matching). **Table A2.** Covariate balance after matching: Number of covariates with STD bias<10%

Declarations

Competing interests
The authors declare they have no competing interests.

Author details
¹Faculty of Education, Beijing Normal University, No. 19, Xinjiekou Wai Street, Haidian District, Beijing 100875, China.
²National Institute of Education Sciences, No. 46, Beisanhuan Zhonglu, Haidian District, Beijing 100088, China. ³School
of Education, Central China Normal University, No. 152, Luoyu Road, Hongshan District, Wuhan 430079, China.

References
Almond, D., and J. Currie. 2011a. Human capital before age five. In *Handbook of labor economics (Vol.4B)*, ed. D. Card and O. Ashenfelter, 1315–1486. Philadelphia, Pa: Elsevier.
Almond, D., and J. Currie. 2011b. Killing me softly: the fetal origin hypothesis. *Journal of Economic Perspectives* 25 (3): 153–172.
Barker, D.J. 1990. The fetal and infant origins of adult disease. *British Medical Journal* 301 (6761): 1111. https://doi.org/10.1136/bmj.301.6761.1111.
Becker, G. 2007. *Human capital: a theoretical and empirical analysis with special reference to education, translated by Guo Hong, Xiong Xiaolin, Wang Xiao, Tan Zhihao*. Beijing: CITIC Press.
Blander, J., P. Gregg, and L. Macmillan. 2007. Accounting for intergenerational income persistence: noncognitive skills, ability and education. *The Economic Journal* 117 (519): C43–C60. https://doi.org/10.1111/j.1468-0297.2007.02034.x.
Blau, D., and J. Currie. 2006. Pre-school, day care, and after-school care: who's minding the kids. In *Handbook of the economics of education (Vol.2)*, ed. E.A. Hanushek and F. Welch, 1163–1278. Amsterdam: Elsevier.
Borraz, F., and A. Cid. 2013. Preschool attendance and school age profiles: a revision. *Children and Youth Services Review* 35 (5): 816–825. https://doi.org/10.1016/j.childyouth.2013.01.023.
Campbell, F.A., and C.T. Ramey. 1995. Cognitive and school outcomes for high risk African American students at middle adolescence: positive effects of early intervention. *American Educational Research Journal* 32 (4): 743–772. https://doi.org/10.3102/00028312032004743.
Chen, Chunjin, and Qian Liu. 2017. A quasi-experiment study of impact of early childhood education on students' academic achievements when they were 15 years old based on the Shanghai data from PISA (in Chinese). *Studies in Early Childhood Education* 1: 3–12.
Cunha, F., and J.J. Heckman. 2007. The technology of skill formation. *American Economic Review* 97 (2): 31–47. https://doi.org/10.1257/aer.97.2.31.
Glewwe, P., Q. Huang, and A. Park. 2017. Cognitive skills, noncognitive skills, and school to work transitions in rural China. *Journal of Economic Behavior and Organization* 134: 141–164. https://doi.org/10.1016/j.jebo.2016.12.009.
Gong, X., D. Xu, and W. Han. 2016. The effects of preschool attendance on adolescent outcomes in rural China. *Early Childhood Research Quarterly* 37: 140–152. https://doi.org/10.1016/j.ecresq.2016.06.003.
Hanushek, E.A. 2010. Developing a skill based agenda for "new human capital research". SSRN Electronic Research Paper (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1889200).
Heckman, J.J. 2006. Skill formation and the economics of investing in disadvantaged children. *Science* 312 (5782): 1900–1902. https://doi.org/10.1126/science.1128898.
Heckman, J.J. 2007. The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences* 104 (33): 13250–13255. https://doi.org/10.1073/pnas.0701362104.
Heckman, J.J. 2013. *Giving kids a fair chance*. Cambridge: The MIT Press.
Heckman, J.J., J. Stixrud, and S. Urzua. 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* 24 (3): 411–482. https://doi.org/10.1086/504455.
Heckman, J.J., R. Pinto, and P. Savelyev. 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103 (6): 2052–2086. https://doi.org/10.1257/aer.103.6.2052.
Hu, Anning. 2015. *The theoretical basis of causal inference in social sciences (in Chinese)*. Beijing: Social Sciences Academic Press.
Huang, G., and Y. Xie. 2013. Cognitive ability. In China people's livelihood development report 2013 (in Chinese), ed. Xie, Y., X. Zhang, J. Li, X. Yu and Q. Ren, 118–147. Beijing: Peking University Press.
Huang, G., Y. Xie, and H. Xu. 2015. Cognitive ability: social correlates and consequences in contemporary China. *Chinese Sociology Review* 47 (4): 287–313. https://doi.org/10.1080/21620555.2015.1032161.
Jiang, Qiuchuan. 2017. The impact of family background and schools' quality on the cognition gap between urban and rural students in China (in Chinese). *Education & Economy* 6: 21–30.
Khor, N., L. Pang, C. Liu, F. Chang, D. Mo, P. Loyalka, and S. Rozelle. 2016. China's looming human capital crisis: upper secondary educational attainment rates and the middle-income trap. *The China Quarterly* 228: 905–926. https://doi.org/10.1017/S0305741016001119.
Li, Minyi. 2011. Improving the national quality of poor children through high quality preschool education: reflection on international experience and our possible policy choices. In *Changing the Odds for Children at Risk, S. B. Neuman, translated by Li Minyi and Huo Liyan, 1-10*. Beijing: Educational Science Publishing House.
Li, Chunling. 2014. The changing trend of educational inequality in China (1940-2010): reexamining the urban-rural gap on educational opportunity (in Chinese). *Sociological Studies* 2: 65–89.
Li, Xiaoman, and Xiangquan Zeng. 2012. Neo human capital theory: research trends of human capital theory based on ability (in Chinese). *Economic Perspectives* 11: 120–126.
Li, H., P. Loyalka, S. Rozelle, B. Wu, and J. Xie. 2015. Unequal access to college in China: how far have poor, rural students been left behind? *The China Quarterly* 221: 185–207. https://doi.org/10.1017/S0305741015000314.
Li, H., P. Loyalka, S. Rozelle, and B. Wu. 2017. Human capital and China's future growth. *Journal of Economic Perspectives* 31 (1): 25–48. https://doi.org/10.1257/jep.31.1.25.
Liu, Jingming. 2014. Ability and background: analysis of mechanisms affecting access to higher education (in Chinese). *Social Sciences in China* 8: 109–128.
Liu, Ji. 2018. Technology disruption and new labor demands: how can education respond (in Chinese). *China Economics of Education Review* 2: 36–51.

Liu, A., and Y. Xie. 2015. Influences of monetary and non-monetary family resources on children's development in verbal ability in China. *Research in Social Stratification and Mobility* 40: 59–70. https://doi.org/10.1016/j.rssm.2015.02.003.

Luo, Renfu, Zhao Qiran, He Min, Liu Chengfang, and Zhang Linxiu. 2009. The empirical analysis of early childhood education of poor rural China (in Chinese). *Studies in Preschool Education* 1: 7–10.

Luo, R., L. Zhang, C. Liu, Q. Zhao, Y. Shi, S. Rozelle, and B. Sharbono. 2012. Behind before they begin: the challenge of early childhood education in rural China. *Australasian Journal of Early Education* 37 (1): 55–64.

Magnuson, K.A., and G.J. Duncan. 2016. Can early childhood interventions decrease inequality of economic opportunity? *The Russell Sage Foundation Journal of the Social Sciences* 2 (2): 123–141. https://doi.org/10.7758/RSF.2016.2.2.05.

Magnuson, K.A., and J. Waldfogel. 2005. Early childhood care and education: effects on ethnic and racial gaps in school readiness. *The Future of Children* 15 (1): 169–196. https://doi.org/10.1353/foc.2005.0005.

Pang, Lijuan, Juan Hu, and Hong Xiumin. 2003. On the value of preschool education (in Chinese). *Studies In Preschool Education* 1: 7–10.

Peng, Junying, Qun Gao, and Xiao Jie. 2011. The outcome of kindergarten education in the western rural area (in Chinese). *Studies in Preschool Education* 7: 9–13.

Rao, N., J. Sun, J. Zhou, and L. Zhang. 2012. Early achievement in rural China: the role of preschool experience. *Early Childhood Research Quarterly* 27 (1): 66–76. https://doi.org/10.1016/j.ecresq.2011.07.001.

Tao, Sha, Liu Hongyun, Zhou Cuimin, Cuicui Wang, Sun Congying, Fen Xu, and Qi Dong. 2015. The roles of school psychological environment in grades 4-6 students' cognitive development: a multilevel analysis of the national representative data (in Chinese). *Journal of Psychological Science* 1: 2–10.

Wang, Huimin, Yuxiao Wu, and Huang Chao. 2017. Family socioeconomic status, preschool education and the adolescents' cognitive & non-cognitive abilities (in Chinese). *Youth Studies* 6: 46–57.

World Bank. 2011. *Early childhood development and education in China: breaking the cycle of poverty and improving future competitiveness*. World Bank Report No.53746-CN.

Wu, Yuxiao. 2013. Educational opportunities of Chinese rural and urban residents in 1978-2008: inequality and evolution (in Chinese). *Social Sciences in China* 3: 4–21.

Xu, H., and Y. Xie. 2015. The causal effects of rural-to-urban migration on children's wellbeing in China. *European Sociological Review* 31 (4): 502–519. https://doi.org/10.1093/esr/jcv009.

Zhang, D., X. Li, and J. Xue. 2015. Education Inequality between rural and urban areas of the People's Republic of China, migrants' children education, and some implications. *Asian Development Review* 32 (1): 196–224. https://doi.org/10.1162/ADEV_a_00042.

Zhao, G., J. Ye, Z. Li, and S. Xue. 2017. How and why do Chinese urban students outperform their rural counterparts? *China Economic Review* 45: 103–123. https://doi.org/10.1016/j.chieco.2017.06.006.

Zhen, Lina, et al. 2016. *Cognitive development and education of preschool children*. Beijing: Beijing Normal University Press.

Publisher's Note