



ARTICLE



<https://doi.org/10.1057/s41599-025-05812-w>

OPEN

# Preschool education and non-cognitive skills: evidence from rural China's junior high students

Wenting Liu<sup>1</sup>, Yijia Li<sup>1</sup>, Lidong Zhang<sup>1</sup>, Yuyang Xie<sup>1</sup>, Jin Zhao<sup>2</sup> & Hongyu Guan<sup>1</sup>✉

Non-cognitive abilities, such as social skills and self-control, are vital for individual success. Preschool education is a critical period for developing these abilities, with long-term benefits often linked to early childhood interventions. Using data from the China Education Panel Survey (CEPS), this study employs OLS regression, PSM, and IV modeling to assess the impact of preschool education on rural junior high school students' non-cognitive abilities. The analysis controls for individual and family characteristics and includes county-fixed effects. The study finds that preschool attendance significantly enhances non-cognitive abilities, especially in dimensions like positive emotions and extraversion. PSM and IV analyses confirm these findings, suggesting that preschool education improves family dynamics and peer relationships, which in turn foster non-cognitive development. The findings highlight the importance of early education investments and suggest that policies supporting preschool education can lead to significant long-term social and economic benefits.

<sup>1</sup>Center for Experimental Economics in Education, Faculty of Education, Shaanxi Normal University, Xi'an, China. <sup>2</sup>School of Marxism, Xi'an University of Finance and Economics, Xi'an, China. ✉email: [hongyuguan0621@gmail.com](mailto:hongyuguan0621@gmail.com)

## Introduction

Non-cognitive abilities, such as creativity, teamwork, and adaptability, are essential for lifelong success (Bowles et al., 2001; Heckman, 2007; Roberts, 2009; Gutman and Schoon, 2013; Duckworth and Yeager, 2015). Research has shown that these skills, alongside cognitive abilities, significantly influence individuals' social status, economic achievement, health, and overall well-being. Early childhood is a critical period for the development of these abilities, and investments in early childhood education, particularly in fostering non-cognitive skills, yield substantial long-term returns (Jencks et al., 1979; Heckman et al., 2006; Edin et al., 2022). Given the significant impact of early education on later life outcomes, it is imperative to thoroughly investigate the long-term effects of preschool education on the development of non-cognitive abilities.

Preschool education plays a multifaceted role in fostering the development of students' non-cognitive abilities (Wang et al., 2024). Through engaging in group play and collaborative activities, students are embedded in social settings that foster positive interactions and essential skills such as sharing, turn-taking, and conflict resolution, all within a supportive and nurturing environment (Ashiabi, 2007; Bodrova and Leong, 2015; Tersi and Matsouka, 2020). Additionally, preschool education emphasizes artistic and creative activities that spark imagination and creativity, while also boosting students' confidence and self-expression (Nagamachi, 2002; Rauf and Tan, 2020; Rainford, 2020). These activities offer a platform for emotional and social intelligence development, enabling students to explore and articulate their thoughts and feelings through diverse creative outlets.

The long-term impact of preschool education on non-cognitive abilities remains inconclusive. Some studies have demonstrated positive effects: a review of 36 preschool programs for low-income students showed improvements in socialization, school enthusiasm, and interpersonal relationships (Barnett, 1995). These programs were also associated with better educational outcomes and long-term behaviors, such as lower rates of teenage pregnancies and criminal activity (Currie and Almond, 2011). The Chicago Longitudinal Study by Reynolds et al. found that enriched preschool environments can enhance both cognitive and non-cognitive skills, with non-cognitive abilities proving particularly influential on criminal behavior, as indicated by reduced incarceration rates (Reynolds et al., 2010).

Conversely, other studies reveal less consistent results regarding the impact of preschool education on non-cognitive development. Apps et al. (2013) observed that, although preschool significantly improves test scores at ages 11, 14, and 16, its impact on non-cognitive aspects is less predictable (Apps et al., 2013). Research by Zhao et al. found no significant association between preschool experience and the development of non-cognitive skills such as social skills and leadership (Zhao and Li, 2024). Additionally, a Canadian study suggested that preschool attendance might negatively affect students' social skills and increase anxiety levels (Baker et al., 2008). Similar findings from

U.S. studies indicate that preschool experiences may, in some cases, diminish learning motivation, self-control, and interpersonal communication skills (Loeb et al., 2007).

Despite the recognized importance of preschool education in nurturing non-cognitive abilities, existing research has three key limitations. First, most studies are concentrated in developed countries, like the United States, and evidence from developing regions remains sparse (Loeb et al., 2007; Baker et al., 2008; Reynolds et al., 2010). Socioeconomic and cultural differences may influence the impact of early education on non-cognitive skills, making it crucial to investigate these relationships in varied contexts. Second, the prevalence of correlational studies limits the

causal inferences we can draw between preschool education and non-cognitive development, particularly given potential endogeneity biases. Lastly, existing research often focuses narrowly on specific non-cognitive skills, such as social interaction or motivation, without employing a holistic framework (Barnett, 1995; Loeb et al., 2007). Thus, there is a need for a comprehensive index grounded in frameworks like the Big Five Personality Model to capture the full spectrum of non-cognitive skills developed in preschool education.

China provides an ideal context for examining the long-term impact of preschool education on non-cognitive skills, particularly given the country's ongoing emphasis on human capital as a driver of economic growth (Li et al., 2005). With rural areas undergoing significant socioeconomic transformations, non-cognitive skills are increasingly valued in both education and the labor market. Leveraging data from the China Education Panel Survey (CEPS), a nationally representative dataset (Zheng et al., 2021; Wang, 2022; Zhang et al., 2023), this study applies instrumental variable analysis to explore the causal relationship between preschool attendance and adolescent non-cognitive development within this distinct environment. This context not only offers insights into human capital formation in rural China but also contributes valuable evidence for other developing countries facing similar challenges.

This study aims to address these research gaps by (1) describing preschool attendance patterns and non-cognitive ability levels among students, (2) assessing the enduring effects of preschool attendance on non-cognitive skills in adolescence, (3) examining heterogeneous effects across different student subgroups, and (4) identifying potential mechanisms through which preschool education impacts non-cognitive development.

## Methods and measures

**Study design and participants.** This study utilizes data from the 2013–2014 China Education Panel Survey (CEPS), a project led by Renmin University of China that tracks Chinese middle school students over the long term to investigate the socioeconomic factors influencing their development. The study employs a phase-stratified probability proportional to size (PPS) sampling method to ensure the sample's representativeness. Initially, 28 counties were selected nationwide, followed by a random selection of schools within these areas, resulting in a thorough investigation of 438 7th and 9th-grade classes and a total of 19,958 students. The survey not only examines students' individual characteristics but also delves into the impact of family, classroom, and school environments on their development, collecting multidimensional information through comprehensive interviews.

In line with the research objectives, the final analytical sample was narrowed to 9443 rural students from 431 classrooms across 112 schools, due to the unavailability of key variables, specifically preschool attendance and non-cognitive ability measures. Further details regarding the sampling methodology, questionnaires, and other aspects are available at <http://ceps.ruc.edu.cn/English/Home.htm>.

**Data collection.** Data collection for our study took place during the second semester of the 2013–2014 academic year, using a comprehensive questionnaire designed to gather detailed information on students' socioeconomic backgrounds, preschool attendance records, non-cognitive ability scores, and various mechanism variables, such as financial and time investments, family dynamics, and peer influence. The survey was distributed to all eligible 7th and 9th-grade students within our sample. In

addition to collecting basic demographic information (e.g., age, gender, and preschool attendance), the survey gathered extensive data on parental characteristics, including education levels and economic status, to provide a nuanced understanding of family background and its potential impact on students' non-cognitive abilities.

### Variables

**Preschool attendance.** The CEPS student survey asked participants if they had attended preschool after turning three, including both You Er Yuan (centers that cater to 50 or more children aged two or three and above) and Xue Qian Ban (offers one year of preschool education prior to formal schooling) (Gong et al., 2016). Ideally, distinguishing between these types of preschools would be crucial for assessing differences in educational quality. However, limitations in the available data prevent such detailed differentiation. As a result, our study uses a binary variable to gauge preschool attendance, signifying whether a child received education before the age of six. A child who attended preschool is assigned a value of 1, and those who did not are assigned a value of 0. This method offers a general measure of the impact of preschool education in China (Zheng et al., 2019; Wang et al., 2023; Zhang and Zhang, 2023).

**Non-cognitive ability score.** In this study, non-cognitive abilities were measured across five dimensions: Conscientiousness, Positive Emotion, Agreeableness, Openness, and Extraversion (Zhang and Zhou, 2023). Each dimension's score was calculated as the mean of all items corresponding to that dimension. The overall non-cognitive ability score for each student was then derived by summing the standardized scores of these five dimensions (Min et al., 2019; Zhao and Chen, 2022). To facilitate comparison, each dimension's score was standardized to a range of 0 to 1 before being summed to create a total non-cognitive ability score.

**Socio-demographic characteristics.** Drawing on theoretical frameworks and empirical evidence (Hampson, 1984; Fabes et al., 1999; Caprara et al., 2014; Magnuson et al., 2016; Schneider and Weber, 2022; He et al., 2023), our analysis includes a comprehensive set of student, parental, and family factors that previous studies have shown to influence students' non-cognitive skills. Recognizing that these factors may act as confounders, particularly concerning our primary variable of interest—preschool attendance—we have incorporated control variables into our empirical models to reduce potential omitted variable bias. Student data includes age, gender, only-child status, left-behind status, and boarding status, while parental data covers parental education levels and family economic status. The variable “Low family economic status” is a binary indicator derived from the question on student family economic conditions, “What is the current economic status of your family?” with responses “difficult,” “medium,” and “wealthy.” We recoded “difficult” as 1, indicating low family economic status, and the other two as 0.

**Statistics analysis.** This research utilized four main statistical approaches. To begin with, we use Ordinary Least Squares (OLS) regression, incorporating the set of covariates previously mentioned, as our initial approach to uncover the link between preschool attendance and the non-cognitive skills of students. This initial step is designed to establish a foundational understanding, which will serve as a benchmark for more advanced analytical methods to be employed later on. Consequently, we start by estimating the following linear equation.

$$Y_i = \beta_0 + \beta_1 \text{Preschool}_i + \beta_2 X_i + \varepsilon_i \quad (1)$$

Where  $Y_i$  represents the standardized scores of non-cognitive abilities for student  $i$ ;  $\text{Preschool}_i$  is a binary indicator for whether student  $i$  is attended preschool, then  $\text{Preschool}_i = 1$  if the student attended; and  $\text{Preschool}_i = 0$  otherwise. The vector  $X_i$  encompasses a range of covariates at both the individual and parental levels, as previously discussed.  $\varepsilon_i$  represents the random error term for each observation. If Eq. (1) is correctly specified, the estimated coefficient for the preschool attendance variable,  $\beta_1$ , will reflect the impact of preschool attendance on the development of non-cognitive abilities. Here,  $i$  indexes each individual observation within our dataset.

Secondly, a significant challenge in the direct application of OLS estimation is the variation in preschool education attendance rates due to disparities in preschool resources among different counties. To accurately account for these differences in preschool education availability and their impact on the development of non-cognitive abilities at the county level, we have incorporated county fixed effects ( $\gamma_i$ ) into our model (Eq. (2)). This adjustment helps to control for both observed and unobserved variations in county-specific characteristics, particularly those related to preschool attendance.

$$Y_i = \beta_0 + \beta_1 \text{Preschool}_i + \beta_2 X_i + \gamma_i + \varepsilon_i \quad (2)$$

Thirdly, the assessment of the long-term impact of preschool education using OLS and county fixed-effect models may be subject to bias due to significant differences between comparison groups. To precisely estimate the long-term effects of preschool education on potential participants, we have adopted the Propensity Score Matching (PSM) method. PSM pairs students in the treatment group with similar ones in the comparison group, attributing differences in non-cognitive abilities to the effects of preschool attendance (Rosenbaum and Rubin, 1985). We apply nearest neighbor and caliper matching with replacement, and bootstrap standard errors over 200 replications. These PSM findings also validate and check for biases in our OLS and county fixed effects model results.

Lastly, to address potential endogeneity, we employ instrumental variable (IV) estimation. *Our selected instruments, the school-level preschool attendance rate and the proportion of kindergartens in the community, are expected to be correlated with the independent variable (preschool attendance) but not with the dependent variable (non-cognitive abilities), thereby serving as valid instruments for this analysis.* Both of our instruments reflect the local supply-side availability of preschool education. These measures influence whether a child attends preschool primarily by increasing accessibility and reducing logistical or financial barriers faced by families. In rural China, preschool enrollment decisions are shaped largely by structural factors such as the presence and capacity of local kindergartens, rather than by individual child characteristics or family preferences related to non-cognitive development. Moreover, the instruments capture conditions specific to the preschool period, which precedes the time frame in which non-cognitive outcomes are measured during junior high school. There is no plausible direct pathway through which the density of kindergartens or school-level enrollment rates several years earlier would independently influence adolescents' non-cognitive skills, except via preschool participation. To further support the exclusion restriction, we conducted placebo analyses using parental occupation as an outcome, which is theoretically unrelated to preschool availability (see in Appendix Table A1). The results showed no significant association with this placebo variable, lending additional credibility to the assumption that their influence on non-cognitive outcomes operates exclusively through preschool attendance. We adjust robust standard errors in parentheses for all regression

**Table 1 Summary statistics of background characteristics.**

Variables, mean (SD)	(1) Full sample	(2) Preschool attendance = 1	(3) Preschool attendance = 0	(4) T-test difference (2)-(3) P value
Dependent variables				
Non-cognition total scores	-0.193 (3.396)	-0.041 (3.383)	-0.680 (3.391)	0.639***
Conscientiousness	0.011 (0.960)	0.018 (0.960)	-0.015 (0.960)	0.033
Neuroticism	-0.014 (0.951)	0.021 (0.944)	-0.128 (0.967)	0.149***
Agreeableness	-0.052 (0.989)	-0.020 (0.981)	-0.153 (1.010)	0.134***
Openness	-0.067 (0.978)	-0.042 (0.974)	-0.146 (0.990)	0.104***
Extraversion	-0.071 (1.003)	-0.019 (0.986)	-0.237 (1.039)	0.219***
Students and family characteristics				
Age (years)	13.989 (1.356)	13.901 (1.309)	14.273 (1.462)	-0.373***
Male (1 = yes 0 = no)	0.511 (0.499)	0.505 (0.500)	0.530 (0.499)	0.025**
Only child (1 = yes 0 = no)	0.266 (0.442)	0.286 (0.452)	0.199 (0.399)	0.087***
Left-behind students (1 = yes 0 = no)	0.261 (0.439)	0.248 (0.432)	0.305 (0.460)	-0.057***
Boarding at school (1 = yes 0 = no)	0.474 (0.499)	0.464 (0.499)	0.505 (0.500)	-0.041***
Father has high school education or above (1 = yes 0 = no)	0.192 (0.394)	0.204 (0.403)	0.155 (0.362)	0.050***
Mother has high school education or above (1 = yes 0 = no)	0.136 (0.343)	0.150 (0.357)	0.094 (0.291)	0.056***
Low family economic status (1 = yes 0 = no)	0.270 (0.444)	0.249 (0.432)	0.340 (0.474)	-0.091***
Observations	9,443	7,189	2,254	

\*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

models, with the analyses conducted using Stata 17.0 (Stata Corp., Texas, USA).

## Results

**Rural students' preschool attendance and non-cognition development.** Analyzing a dataset of 9443 students, we discovered that 76.13% had preschool experience, correlating with various demographic factors. Table 1 reveals a balanced gender distribution in the sample, with 51.1% ( $n = 4825$ ) male students and 48.9% ( $n = 4618$ ) female students. The average age of participants was 13.99 years, with a standard deviation of 1.36 years. In terms of family structure, 26.6% were from only-child families, 26.1% were left-behind students, and 47.4% were boarding students. Regarding family economics, about 27% of students were from the poorest households. Lastly, only 19.2% of fathers and 13.6% of mothers had education levels beyond high school. Data also shows that rural preschool attendees typically come from wealthier backgrounds and are younger, only child, non-left-behind, and non-boarding, significantly different from non-attendees ( $P < 0.001$ ). Their parents also have higher education and income levels ( $P < 0.001$ ). These differences emphasize the need to account for baseline disparities when evaluating the impact of preschool on non-cognitive skill development in rural China.

The data further indicate that individuals with preschool experience tend to have higher non-cognitive ability levels than those without such experience (mean  $\pm$  SD:  $-0.041 \pm 3.383$  VS  $-0.680 \pm 3.391$ ). This suggests a possible positive impact of preschool education on the development of non-cognitive abilities. The data analysis reveals significant differences among five non-cognitive ability dimensions: positive emotions (mean  $\pm$  SD:  $0.021 \pm 0.944$  VS  $-0.128 \pm 0.967$ ), agreeableness (mean  $\pm$  SD:  $-0.020 \pm 0.981$  VS  $-0.153 \pm 1.010$ ), openness (mean  $\pm$  SD:  $-0.042 \pm 0.974$  VS  $-0.146 \pm 0.990$ ), and extraversion (mean  $\pm$  SD:  $-0.019 \pm 0.986$  VS  $-0.237 \pm 1.039$ ). Junior high school students who had received preschool education significantly outperformed those who had not. However, there seems to be little difference in the index of conscientiousness between individuals with and without preschool attendance experience (mean  $\pm$  SD:  $0.018 \pm 0.960$  VS  $-0.015 \pm 0.960$ ).

**Long-term effects of the preschool attendance on rural students' non-cognition development.** Table 2 details the long-term effects of preschool attendance on rural students' non-cognitive abilities, with columns 1–6 adjusting for individual and family factors, and columns 7–12 adding county-fixed effects. The key figures are the coefficients for preschool attendance, with robust standard errors shown in parentheses. Our empirical study on the

**Table 2 The effect of preschool attendance on student's non-cognition development.**

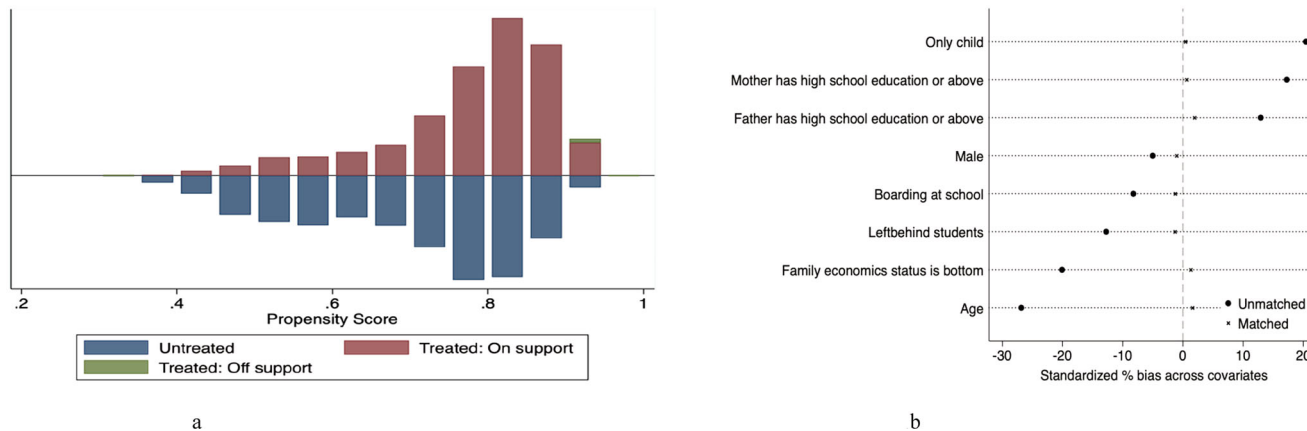
Variables	Full sample											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Non-cognition total scores	0.386***	-0.021	0.095***	0.086***	0.055**	0.171***	0.302***	-0.018	0.073***	0.081***	0.043*	0.124***
Preschool attendance (1 = yes 0 = no)	(0.082)	(0.023)	(0.023)	(0.024)	(0.024)	(0.025)	(0.083)	(0.023)	(0.024)	(0.025)	(0.024)	(0.025)
Age (years)	-0.232***	-0.116***	-0.063***	-0.019**	-0.020***	-0.013*	-0.164***	-0.124***	-0.068***	0.006	0.002	0.020**
Male (1 = yes 0 = no)	(0.026)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.027)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Only child (1 = yes 0 = no)	-0.439***	-0.245***	0.061***	-0.150***	-0.028	-0.077***	-0.362***	-0.232***	0.067***	-0.133***	-0.007	-0.056***
Left-behind students (1 = yes 0 = no)	(0.069)	(0.019)	(0.019)	(0.020)	(0.020)	(0.021)	(0.067)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)
Boarding at school (1 = yes 0 = no)	0.242***	0.011	0.029	0.053**	0.075***	0.074***	-0.108	-0.022	0.007	-0.032	-0.017	-0.043*
Father has high school education or above (1 = yes 0 = no)	(0.082)	(0.024)	(0.023)	(0.024)	(0.024)	(0.024)	(0.086)	(0.025)	(0.025)	(0.026)	(0.025)	(0.025)
Mother has high school education or above (1 = yes 0 = no)	-0.641***	-0.067***	-0.155***	-0.135***	-0.126***	-0.158***	-0.403***	-0.061***	-0.123***	-0.090***	-0.078***	-0.051**
Low family economic status (1 = yes 0 = no)	(0.079)	(0.022)	(0.022)	(0.024)	(0.023)	(0.024)	(0.083)	(0.024)	(0.024)	(0.025)	(0.024)	(0.025)
Fixed at the county level	-0.039	0.029	0.007	-0.002	0.008	-0.082***	0.060	0.067**	0.009	-0.013	0.003	-0.007
Constant	(0.071)	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.095)	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)
R <sup>2</sup>	0.247***	-0.016	0.022	0.094***	0.106***	0.041	0.279***	-0.006	0.033	0.091***	0.109***	0.053*
Observations	(0.094)	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)	(0.094)	(0.028)	(0.027)	(0.027)	(0.027)	(0.028)
	0.459***	0.032	0.089***	0.084***	0.119***	0.135***	0.340***	0.035	0.065**	0.047	0.080**	0.113***
	(0.111)	(0.033)	(0.031)	(0.033)	(0.032)	(0.033)	(0.112)	(0.033)	(0.032)	(0.033)	(0.032)	(0.033)
	-0.648***	-0.006	-0.170***	-0.167***	-0.170***	-0.135***	-0.434***	0.011	-0.138***	-0.119***	-0.121***	-0.066***
	(0.078)	(0.022)	(0.022)	(0.023)	(0.023)	(0.024)	(0.079)	(0.022)	(0.023)	(0.024)	(0.023)	(0.024)
	NO	NO	NO	NO	NO	NO	Yes	Yes	Yes	Yes	Yes	Yes
	3.168***	1.779***	0.827***	0.267**	0.201*	0.094	3.602***	1.908***	0.988***	0.281**	0.229	0.196
	(0.378)	(0.106)	(0.107)	(0.112)	(0.110)	(0.113)	(0.484)	(0.139)	(0.143)	(0.142)	(0.141)	(0.138)
	0.043	0.045	0.029	0.025	0.022	0.029	0.088	0.070	0.043	0.060	0.060	0.082
	9443	9443	9443	9443	9443	9443	9443	9443	9,443	9,443	9,443	9,443

Each column corresponds to a separate regression. Covariates include age, only child, left behind students, boarding at school, father has high school education or above, mother has high school education or above, family economics status. Robust standard errors in parentheses, fixed at the county level. \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table 3 PSM results of the effect of preschool attendance on student's non-cognition development.**

Dependent variable	Full sample				
	ATT	Std. Err	z	P value	Observations
PSM method: nearest neighbor estimator					
Non-cognition total scores	0.336***	0.107	3.13	0.002	9443
Conscientiousness	-0.027	0.033	-0.83	0.407	9443
Neuroticism	0.086**	0.033	2.57	0.010	9443
Agreeableness	0.089**	0.039	2.23	0.026	9443
Openness	0.062*	0.034	1.8	0.072	9443
Extraversion	0.127***	0.036	3.79	<0.001	9443
PSM method: caliper estimator					
Non-cognition total scores	0.297***	0.089	3.31	0.001	9443
Conscientiousness	-0.015	0.029	-0.55	0.582	9443
Neuroticism	0.079**	0.029	2.71	0.007	9443
Agreeableness	0.089**	0.026	3.47	0.001	9443
Openness	0.049*	0.027	1.85	0.065	9443
Extraversion	0.095***	0.025	3.79	<0.001	9443

Each column corresponds to a separate regression. Covariates include age, only child, left behind students, boarding at school, father has high school education or above, mother has high school education or above, family economics status. Robust standard errors in parentheses, fixed at the county level, clustered at the school level. \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.



**Fig. 1 Diagnostics for propensity score matching.** **a** Overlap in the support of the covariates after between the attendees and non-attendees. **b** Standardized mean differences of baseline covariates before and matching (Love plot).

long-term impact of preschool attendance on non-cognitive abilities found two main results. The regression model, controlling for individual and family factors (Table 2, columns 1–6), shows a significant positive effect of preschool on non-cognitive abilities ( $=0.386, P < 0.01$ ). It also positively influences positive emotions ( $=0.095, P < 0.01$ ), agreeableness ( $=0.086, P < 0.01$ ), openness ( $=0.055, P < 0.05$ ), and extraversion ( $=0.171, P < 0.01$ ), but not conscientiousness. Columns 7–12 of Table 2 confirm the robustness of these findings after accounting for county fixed effects. *Our analysis also identifies several key factors that enhance students' non-cognitive abilities and their corresponding dimensions. Notably, younger students, females, only children, non-left-behind students, those with highly educated parents, and those from families with better economic status tend to have higher non-cognitive scores ( $P < 0.01$ ).*

In addition, we performed Oster's  $\delta$  sensitivity analysis for the OLS estimates, as shown in Appendix Table A2. The  $\delta$  value for reducing the estimated effect to zero is 1.074, suggesting that unobserved confounders would need to be at least as influential as the observed covariates to nullify our findings. While the  $\delta$  value drops to 0.544, indicating a reduction in effect by half, suggesting some sensitivity to moderate unobserved bias, our key conclusions remain robust.

**Propensity score matching (PSM).** Table 3 presents the propensity score matching (PSM)-adjusted outcomes for non-cognitive abilities. Rows 1 and 7, column 1 show regressions for total non-cognitive scores, while rows 2–6 and 8–12, column 1 detail the long-term effects of preschool attendance on five non-cognitive dimensions. Specifically, preschool attendees show 7.9–8.6 percentage points higher positive emotions, 8.9 points higher agreeableness, 4.9–6.2 points higher openness, and 9.5–12.7 points higher extraversion compared to non-attendees. The PSM results suggest that the results are robust after considering the selection of observations.

Figure 1a depicts the post-matching probability densities for preschool attendees and non-attendees using the county-fixed effects model, showing good overlap between the groups. This substantial overlap, along with a balanced covariate distribution, validates the subsequent analysis on the matched sample of 9443. *To further assess the robustness of our PSM results to potential unobserved confounders, we report standardized mean differences (SMDs) before and after matching (Tan, Cai, and Bodovski, 2022), and present a Love plot to visually display the improvement in covariate balance. Results show that no covariate exhibits an SMD above 10% after matching, suggesting satisfactory balance across treatment groups (Details in Table 4, Table A3, and Fig. 1b). We*

conducted a Rosenbaum bounds sensitivity analysis to assess the robustness of the estimated treatment effect to potential hidden bias. The results (see in Appendix Table A4) show that the findings remain statistically significant even under a sensitivity parameter  $\Gamma = 2.0$ , suggesting that an unobserved confounder would have to increase the odds of receiving treatment by 100% to invalidate the results. This indicates strong robustness to unobserved selection bias.

**IV estimates.** Table 5 presents the results from the two-stage least squares (2SLS) estimation using both the school-level preschool enrollment rate and the community kindergarten density as instruments for preschool attendance. Including both instruments jointly in the first stage enhances the relevance of the instruments and allows for overidentification tests to assess instrument validity.

The first-stage F-statistic is 22.82, exceeding the conventional threshold of 10 and suggesting no weak instrument problem. The

Kleibergen–Paap rk LM statistic yields a  $p$  value of 0.0007, rejecting the null hypothesis of underidentification, confirming that the model is well-identified. The Hansen J-test produces  $p$ -values above 0.30 across all specifications (e.g., 0.550 in column 2), indicating that the instruments are not correlated with the second-stage error term and thus satisfy the exclusion restriction.

In the second stage, preschool attendance is positively and significantly associated with non-cognitive outcomes. Specifically, attending preschool increases the total non-cognitive score by 4.396 points (standard error = 2.381,  $p < 0.10$ ). Among subdimensions of non-cognitive traits, significant positive effects are found for Neuroticism (coefficient = 0.753, SE = 0.400,  $p < 0.10$ ), Agreeableness (coefficient = 1.321, SE = 0.654,  $p < 0.05$ ), and Openness (coefficient = 1.126, SE = 0.558,  $p < 0.05$ ). The coefficients for Conscientiousness (0.516, SE = 0.533) and Extraversion (0.680, SE = 0.546) are positive but not statistically significant at conventional levels. These results suggest that preschool participation has favorable impacts on overall non-cognitive abilities and several specific personality traits, even after accounting for potential endogeneity.

**Robustness check.** We perform multiple robustness checks to verify the stability of our basic model’s findings, with the outcomes presented in Table 6. Given the positive impact of preschool education on child development, it is reasonable to expect that extended exposure to quality preschool would yield superior outcomes. To explore this, we categorized preschool duration into four groups: less than 1 year (duration  $\leq 1$  year), 1–2 years (1 year  $<$  duration  $\leq 2$  years), and above 2 years (duration  $> 2$  years). We constructed three dummy variables for these groups, using children with no preschool experience as the reference category. Our findings, as detailed in Table 6, Panel A, reveal that students with the longest preschool tenure, particularly those attending for over 2 years, experienced the most substantial benefits in terms of non-cognitive development.

We investigated the significance of the timing of preschool entry (Table 6, Panel B). We categorized preschool starting age into five groups: under 2 years (age  $< 2$ ), 2–3 years ( $2 \leq$  age  $< 3$ ), 3–4 years ( $3 \leq$  age  $< 4$ ), 4–5 years ( $4 \leq$  age  $< 5$ ), and over 5 years (age  $\geq 5$ ), with an additional control group for those who did not

**Table 4 Summary of standard bias (in %) before and after matching.**

Preschool attendees vs. non-attendees	Sample	
	Unmatched	Matched
%Variance	100	100
Mean bias	15.4	1.2
B	41.0*	3.7
R	0.89	0.96
Number of predictors in the models	8	8
Number of predictors with an S.B. greater than 5	10	0
Number of predictors with an S.B. between 1.01 and 1.99	0	5
Number of predictors with an S.B. less than 1.0	0	3
Number of predictors with a $p$ value less than 0.001	6	0

\* indicate statistical significance at the 10% level

**Table 5 IV estimate.**

Variables	(1) Preschool attendance	(2) Non-cognition total scores	(3) Conscientiousness	(4) Neuroticism	(5) Agreeableness	(6) Openness	(7) Extraversion
Preschool attendance at the school level	0.463*** (0.086)						
Is there a preschool in your community	0.026*** (0.009)						
Preschool attendance		4.396* (2.381)	0.516 (0.533)	0.753* (0.400)	1.321** (0.654)	1.126** (0.558)	0.680 (0.546)
Test for under identifying restrictions ( $p$ value)	0.0007						
F test for weak instruments	22.82						
Test for over identifying restrictions ( $p$ value)		0.5500	0.3076	0.9302	0.9745	0.8895	0.3498
Constant		-1.041 (3.071)	1.271* (0.690)	0.120 (0.533)	-1.098 (0.837)	-0.914 (0.727)	-0.420 (0.719)
$R^2$		-0.162	0.015	-0.042	-0.214	-0.152	0.029
Observations		7980	7980	7980	7980	7980	7980

Each column corresponds to a separate regression. Covariates include age, only child, left behind students, boarding at school, father has high school education or above, mother has high school education or above, family economics status. Robust standard errors in parentheses, fixed at the county level, clustered at the school level. The duration of preschool attendance and the start time of preschool attendance variables have some missing values. \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table 6 Robustness check of the effect of preschool attendance on student's non-cognition development.**

Variables	Full sample					
	(1) Non-cognition total scores	(2) Conscientiousness	(3) Neuroticism	(4) Agreeableness	(5) Openness	(6) Extraversion
Panel A: The duration of preschool attendance						
Less than 1 year (1 = yes 0 = no)	0.181 (0.115)	-0.014 (0.033)	0.070** (0.032)	0.050 (0.035)	-0.011 (0.035)	0.087** (0.035)
For 1-2 years (1 = yes 0 = no)	0.260** (0.106)	-0.009 (0.030)	0.085*** (0.031)	0.065** (0.031)	0.019 (0.031)	0.100*** (0.032)
For 2 years and above (1 = yes 0 = no)	0.442*** (0.095)	-0.007 (0.027)	0.077*** (0.027)	0.121*** (0.028)	0.097*** (0.028)	0.154*** (0.028)
Observations	8893	8893	8893	8893	8893	8893
Panel B: The start time of preschool attendance						
Younger than 2 years	0.038 (0.908)	0.062 (0.200)	0.072 (0.201)	-0.156 (0.285)	-0.125 (0.223)	0.185 (0.204)
Between 2 and 3 years (including 2 years)	0.436** (0.220)	-0.093 (0.065)	-0.027 (0.065)	0.131** (0.064)	0.125** (0.063)	0.300*** (0.064)
Between 3 and 4 years (including 3 years)	0.528*** (0.106)	-0.013 (0.030)	0.089*** (0.031)	0.142*** (0.031)	0.125*** (0.031)	0.186*** (0.032)
Between 4 and 5 years (including 4 years)	0.353*** (0.101)	-0.005 (0.029)	0.089*** (0.029)	0.102*** (0.030)	0.057* (0.029)	0.110*** (0.030)
Older than 5 years (including 5 years)	0.151 (0.102)	-0.005 (0.029)	0.070** (0.029)	0.033 (0.031)	-0.021 (0.030)	0.075** (0.031)
Observations	8928	8928	8928	8928	8928	8928
Panel C: Random classes						
	0.350*** (0.110)	-0.007 (0.031)	0.088*** (0.032)	0.074** (0.033)	0.041 (0.033)	0.154*** (0.032)
Observations	5689	5689	5689	5689	5689	5689
Panel D: First born						
	0.251* (0.137)	-0.032 (0.038)	0.040 (0.039)	0.068* (0.041)	0.015 (0.041)	0.162*** (0.042)
Observations	3270	3270	3270	3270	3270	3270

Each column corresponds to a separate regression. Covariates include age, only child, left behind students, boarding at school, father has high school education or above, mother has high school education or above, family economics status. Robust standard errors in parentheses, fixed at the county level, clustered at the school level. The duration of preschool attendance and the start time of preschool attendance variables have some missing values. \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

*attend preschool.* Considering that students starting earlier might have the longest exposure, we controlled for the duration of preschool in our model to isolate timing effects from dosage effects. Our results indicate that students who began preschool between 3 and 4 years of age (including 3 years) saw the greatest benefits for their non-cognitive development.

Furthermore, the lack of random class assignment could unbalance the sample, leading to potential bias. To counter this, regressions in Table 5, Panel C, utilized a randomly sorted sample, yielding coefficients for non-cognitive skills and their five dimensions (=0.074–0.350,  $P < 0.05$ ). Concerning birth order, variations in results may stem from differences in lower-parity students or parental investment choices varying with birth order, as noted by Black, Devereux, Salvanes (2005) (Black et al., 2005). Additionally, having older siblings in preschool can help bypass waiting lists. The last panel (D) of Table 6 presents findings specific to first-born students, with coefficient estimates aligning with our previous results.

**Heterogeneity analysis.** To examine the differential impact of preschool attendance on students' non-cognitive abilities, we

conducted a heterogeneity analysis from eight distinct perspectives, as presented in Table 7. These include age, gender, only-child status, left-behind status, boarding status, father's educational level (high school or above), mother's educational level (high school or above), and family economic status (poorest). Our findings indicate that certain subgroups exhibit a stronger effect from preschool attendance. Specifically, younger students, male students, non-only students, non-left-behind students, boarding students, and students with parents who have lower education levels or better family economic status tend to experience more pronounced gains in non-cognitive abilities.

**Mechanism.** To understand the pathways through which preschool attendance influences students' non-cognitive abilities, we focus on two key dimensions: family resource investment and peer interactions. These dimensions were selected because they represent fundamental aspects of a child's early social and emotional environment that are likely to shape non-cognitive skills. Family resource investment involves the time, money, and attention parents provide to support their child's development, fostering skills such as perseverance, emotional regulation, and

**Table 7 Heterogeneity analysis of the effect of preschool attendance on student's non-cognition development.**

Panel A: variables	Ages		Gender		Only child		Left-behind students	
	≥13 years	<13 years	Male	Female	Yes	No	Yes	No
Preschool attendance (1 = yes 0 = no)	0.312*** (0.089)	0.524** (0.223)	0.301** (0.120)	0.291*** (0.113)	0.272 (0.188)	0.303*** (0.092)	0.124 (0.151)	0.382*** (0.098)
Students characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.292*** (0.289)	1.800 (1.902)	3.349*** (0.704)	3.370*** (0.660)	5.452*** (0.933)	2.871*** (0.578)	1.016 (1.205)	4.174*** (0.547)
R <sup>2</sup>	0.084	0.112	0.066	0.117	0.085	0.093	0.085	0.086
Observations	8018	1425	4826	4617	2508	6935	2469	6974

Panel B: variables	Boarding at school		Father has high school education or above		Mother has high school education or above		Family economics status is bottom	
	Yes	No	Yes	No	Yes	No	Yes	No
Housework participation (1 = yes 0 = no)	0.364*** (0.114)	0.229* (0.119)	0.041 (0.204)	0.356*** (0.091)	0.158 (0.268)	0.316*** (0.087)	0.278* (0.150)	0.317*** (0.099)
Students characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	2.935 (3.128)	4.887*** (0.632)	5.296*** (1.052)	3.249*** (0.549)	4.105*** (1.334)	3.442*** (0.524)	1.842* (1.055)	4.083*** (0.551)
R <sup>2</sup>	0.060	0.117	0.092	0.088	0.076	0.088	0.075	0.083
Observations	4476	4967	1818	7625	1288	8155	2554	6889

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table 8 Mechanism analysis of the effect of preschool attendance on student's non-cognition development.**

Variables	Full sample	
	Estimate	Observations
Panel A: monetary, time investment, and family dynamics		
A lot of books at home	0.105***	9428
Exercise with students	0.089*	8827
Visit museums, zoos, and science museums with parents	0.103***	9361
Going out to the movies, shows, and sports games with parents	0.090***	9388
Parent-students relationship	0.037***	9384
Talk about school things with students	0.062***	8695
Talk about student's mood with students	0.042**	8651
Panel B: peer level		
Visit museums, zoos, and science museums with classmates	0.077***	9223
Going out to the movies, shows, and sports games with classmates	0.101***	9195

Each column corresponds to a separate regression. Covariates include age, only child, left behind students, boarding at school, father has high school education or above, mother has high school education or above, and family economics status. Robust standard errors in parentheses, fixed at the county level, clustered at the school level. These mechanism variables have some missing values. \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

social communication. Peer interactions, on the other hand, expose students to diverse social scenarios, enhancing abilities like empathy, cooperation, and adaptability, which are essential for long-term social success. Table 8 presents the detailed findings.

*Monetary, time investment, and family dynamics.* To examine how preschool attendance fosters non-cognitive skills through family resources, we analyzed a range of related activities, which include both material resources and parent-child interactions. Specifically, nine channels were examined to assess the effect of preschool attendance on students' non-cognitive development: a large collection of books at home, exercising with parents, visiting cultural venues (such as museums, zoos, or science museums) with parents, attending movies, shows, or sports games with parents, quality of the parent-child relationship, discussing school

matters with parents, talking about the child's emotional experiences with parents, visiting cultural venues with classmates, and attending movies, shows, or sports games with classmates.

In Panel A of Table 8, we report how the experience of preschool attendance positively affects indicators of family resource investment. The results show statistically significant effects of preschool attendance on indicators including: "A large collection of books at home (=0.105,  $P < 0.01$ )", "Exercising with parents (=0.089,  $P < 0.1$ )", "Visiting cultural venues with parents (=0.103,  $P < 0.01$ )", "Attending entertainment events with parents (=0.090,  $P < 0.01$ )", "Parent-child relationship quality (=0.037,  $P < 0.01$ )", "Discussing school matters with parents (=0.062,  $P < 0.01$ )", and "Talking about the child's emotions with parents (=0.042,  $P < 0.01$ )". These findings suggest that increased family time and resource investment fostered by preschool

attendance may play a role in developing student's non-cognitive skills, enhancing their social, emotional, and cognitive resilience.

*Peer level.* Research by Li and Zhao (2022) suggests that peer interactions are instrumental in improving students' non-cognitive skills (Li and Zhao, 2022). In line with this, our study used two indicators to evaluate the impact of preschool attendance on peer relationships: "visiting cultural venues with classmates" and "attending entertainment events with classmates". As shown in the results (Table 7, Panel B), both indicators had significantly positive coefficients, implying that preschool attendance enhances student's non-cognitive abilities by promoting positive and frequent interactions with peers.

In addition to the baseline mechanism analysis, we performed KHB decomposition analyses. The results indicate that family monetary and time investments, as well as peer interactions, partially explain the positive effect of preschool attendance on non-cognitive outcomes, though significant direct effects remain (see Appendix Table A5).

### Conclusion and discussions

With the labor market increasingly valuing human capital, the role of non-cognitive skills is gaining importance. Preschool education significantly contributes to the development of these skills by providing an early foundation in essential abilities, such as adaptability, teamwork, and social engagement. This study investigates the long-term effects of preschool attendance on non-cognitive development among junior high school students in rural China, utilizing nationally representative data from the 2013–2014 China Education Panel Survey (CEPS). Specifically, it examines how preschool attendance affects factors tied to students' development, including family resource investment, time investment, and peer interactions. Our findings contribute new insights into the ongoing benefits of preschool education within rural contexts, highlighting its potential to enhance students' non-cognitive abilities in meaningful ways.

The results demonstrate that students with preschool experience exhibit stronger development in non-cognitive dimensions, particularly positive emotions, agreeableness, openness, and extraversion. Furthermore, the data reveal that certain demographic factors—such as being an only child, living with parents, not boarding, having more educated parents, and coming from economically advantaged backgrounds—are associated with higher rates of preschool attendance in China. This suggests that access to preschool is partially shaped by socioeconomic status, raising important considerations for equitable early education access. This conclusion is consistent with other studies (Wu, 2011; Zhang, 2013; Gong et al., 2016). The observed long-term benefits of preschool attendance for non-cognitive development highlight the importance of expanding preschool access, especially in under-resourced rural areas, where these skills are crucial for personal and academic growth in an evolving economic landscape.

Using four identification strategies—OLS regression, county fixed effects, propensity score matching, and instrumental variables—we discovered a significant positive long-term impact of preschool attendance on students' non-cognitive abilities. These results align with Chinese studies on long-term effects (Gong et al., 2016) as well as findings from international samples (Nores and Barnett, 2010). Our data indicate that longer preschool attendance (two years or more) confers greater benefits in non-cognitive abilities and that enrolling between the ages of 3 and 4 yields the most substantial advantages for non-cognitive development. Thus, our findings underscore the critical role of preschool in fostering long-term non-cognitive abilities in rural China.

Our heterogeneity analysis offers further insights into the nuanced effects of preschool attendance across different student profiles. Firstly, the impact of preschool attendance on non-cognitive development is more pronounced for non-Only-Child students, possibly due to a lack of sibling interaction in only-child households that might otherwise foster social skills (Falbo and Poston Jr, 1993; Tobin et al., 2009). Secondly, left-behind students—those separated from one or both parents due to migration—appear less able to benefit from preschool experiences, likely due to a combination of both reinforcing and offsetting social effects. Finally, younger students, boys, non-Only students, boarders, and those with parents with lower educational levels show the greatest benefits from preschool attendance. These findings suggest the need for policies that alleviate financial burdens on vulnerable households and provide affordable preschool access to rural families. Such support can amplify the benefits of preschool, reduce developmental disparities, and foster greater educational equity between disadvantaged and advantaged students.

We identify two key mechanisms by which preschool education could enhance students' non-cognitive skills: monetary and time investment from families, as well as enriched peer interactions, both of which contribute positively to non-cognitive development. Preschool attendance frees up parents' time (Hirshberg et al., 2005; Barnett and Jung, 2021), which they can use for work, potentially raising family income for educational resources (Hirshberg et al., 2005; Guo et al., 2024), or for rest, improving their capacity for supportive parenting (Wu, 2011; Havnes and Mogstad, 2015). Additionally, preschool may expose parents to child development information, supporting more positive and engaged parenting practices (Wu, 2011; Ladd, 2016), positively influencing family dynamics. On the peer side, preschool provides a structured environment where students can regularly interact with peers, helping them develop interpersonal skills like cooperation, conflict resolution, and empathy (Boivin et al., 2005; Eivers et al., 2012; Cappelen et al., 2020; Xiao et al., 2022). These early peer interactions allow students to practice social behaviors and adaptability in a safe setting, building confidence and social awareness (Repper and Carter, 2011; Xiao et al., 2022). Together, these enhanced family and peer experiences offer a foundation for developing essential non-cognitive skills that support long-term success.

We recognize three limitations in our study that merit consideration. Firstly, due to data limitations, we were unable to analyze the effects of specific preschool types or investigate the mechanisms in depth, despite offering speculative explanations. Future research with more detailed data is necessary to gain a clearer understanding of these nuances. Secondly, although we employed four different strategies to address endogeneity concerns, unobserved confounders may still influence our findings. Further investigation with stronger identification methods could help clarify the link between preschool attendance and non-cognitive outcomes. Lastly, our use of cross-sectional data limits our ability to establish causality definitively. Future studies employing longitudinal panel data would provide a more robust basis for causal inferences.

Despite these limitations, our study makes an important contribution by offering a comprehensive analysis of how preschool attendance affects the development of non-cognitive skills in students. The findings have significant policy implications for student development in rural China, suggesting that expanding preschool access could be an effective way to enhance non-cognitive outcomes. Given the limited preschool availability for rural students, as highlighted by Gong et al. (2016) (Gong et al., 2016), efforts to increase access for

students from rural households are crucial. Expanding pre-school opportunities in rural areas could help bridge developmental disparities between rural and urban regions and promote equity from early childhood onward within rural communities.

Future research could build upon this study by exploring the moderating effects of factors such as family engagement, community support, and regional educational policies would offer deeper insights into how preschool interventions can be tailored to different socioeconomic contexts. Further studies could also investigate the longitudinal impact of preschool education on various life outcomes, thereby enhancing our understanding of the role of early childhood interventions across the lifespan.

**Data availability**

Database available from the China Education Panel Survey (CEPS) repository, <http://ceps.ruc.edu.cn/English/Home.htm>.

Received: 1 December 2024; Accepted: 20 August 2025;

Published online: 29 September 2025

**Appendix**

Tables 9–13

**Table 9 Placebo checks for IV estimate.**

Variables	(1) Preschool attendance	(2) Parent vocation
Preschool attendance at the school level	0.470*** (0.087)	
Is there a preschool in your community	0.026*** (0.009)	
Preschool attendance		0.509 0.988
Test for under-identifying restrictions (p value)	0.0006	
F test for weak instruments	23.09	
Test for over-identifying restrictions (p value)		0.6402
Constant		5.924 (1.328)
R <sup>2</sup>		0.114
Observations		7923

Each column corresponds to a separate regression. Covariates include age, only child, left behind students, boarding at school, father has high school education or above, mother has high school education or above, family economics status. Robust standard errors in parentheses, fixed at the county level, clustered at the school level. The duration of preschool attendance and the start time of preschool attendance variables have some missing values. \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table 10 Oster's  $\delta$  tests.**

	R_max = 0.095
beta = 0 (zero effect)	1.074
beta = 0.193(half effect)	0.544

**Table 11 Summary statistics of background characteristics after matching.**

Variables, mean	(1) Preschool attendance = 1	(2) Preschool attendance = 0	(3) %bias	(4) T-test Difference (1)-(2) P value	(5) V(T)/V(C)
Age (Year)	13.902	13.88	1.6	0.98	0.328
Male, 1 = yes 0 = no	0.506	0.511	-1.0	-0.63	0.530
Only child, 1 = yes 0 = no	0.285	0.283	0.4	0.24	0.810
Left-behind students, 1 = yes 0 = no	0.248	0.254	-1.3	-0.82	0.414
Boarding at school, 1 = yes 0 = no	0.465	0.471	-1.3	-0.77	0.442
Father has high school education or above, 1 = yes 0 = no	0.203	0.196	1.9	1.11	0.268
Mother has high school education or above, 1 = yes 0 = no	0.148	0.146	0.6	0.34	0.731
Low family economic status 1 = yes 0 = no	0.249	0.243	1.3	0.82	0.414
Observations	7176	2254			

\*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Table 12 Rosenbaum bounding analysis for sensitivity test.**

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0	0	1	1	1	1
1.1	0	0	1	1	1	1
1.2	0	0	1	1	1	1
1.3	0	0	1	1	1	1
1.4	0	0	0.5	1	0.5	1
1.5	0	0	0.5	1	0.5	1
1.6	0	0	0.5	1	0.5	1
1.7	0	0	0.5	1	0.5	1
1.8	0	0	0.5	1	0.5	1
1.9	0	0	0.5	1	0.5	1
2	0	0	0.5	1	0.5	1

Gamma: log odds of differential assignment due to unobserved factors.  
 sig+ :upper bound significance level.  
 sig- :lower bound significance level.  
 t-hat+ :upper bound Hodges-Lehmann point estimate.  
 t-hat- :lower bound Hodges-Lehmann point estimate.  
 CI+ :upper bound confidence interval ( $\alpha = 0.95$ ).  
 CI- :lower bound confidence interval ( $\alpha = 0.95$ ).

**Table 13 KHB analysis for mediation effect.**

	Monetary, time investment and family dynamics					Peer level			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variables	A lot of books at home	Exercise with students	Visit museums, zoos, science museums with parents	Going out to the movies, shows, sports games with parents	Parent-students relationship	Talk about school things with students	Talk about student's mood with students	Visit museums, zoos, science museums with classmates	Going out to the movies, shows, sports games with classmates
Reduced model	0.382*** (0.079)	0.358*** (0.083)	0.391*** (0.081)	0.384*** (0.080)	0.382*** (0.078)	0.358*** (0.082)	0.360*** (0.081)	0.353*** (0.081)	0.357*** (0.081)
Full model	0.243*** (0.080)	0.299*** (0.083)	0.303*** (0.081)	0.305*** (0.080)	0.335*** (0.078)	0.264*** (0.082)	0.321*** (0.081)	0.260*** (0.081)	0.276*** (0.082)
Diff	0.139*** (0.020)	0.059*** (0.015)	0.088*** (0.015)	0.078*** (0.016)	0.047** (0.023)	0.094*** (0.020)	0.039* (0.023)	0.093*** (0.016)	0.081*** (0.015)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	9428	8827	9361	9388	9384	8695	8651	9223	9195

This table reports the results of the Karlson-Holm-Breen (KHB) decomposition analysis to examine how family investments and peer interactions mediate the relationship between preschool attendance and non-cognitive outcomes. The dependent variable in all models is non-cognitive outcomes. The reduced models estimate the total effect of preschool attendance without including mediators, while the full models include each mediator separately. The "Diff" rows show the indirect effects, representing the portion of the total effect transmitted through each mediator. All mediators show statistically significant indirect effects, with the largest reductions observed for "A lot of books at home" (indirect effect = 0.139,  $p < 0.01$ ) and "Visit museums, zoos, science museums with parents" (indirect effect = 0.088,  $p < 0.01$ ). These results indicate that both monetary and time investments by families, as well as peer-related activities, partially mediate the positive effect of preschool attendance on non-cognitive development. However, significant direct effects remain, suggesting that preschool has additional benefits beyond these family and peer mechanisms. Standard errors are in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

**References**

Apps P, Mendolia S, Walker I (2013) The impact of pre-school on adolescents' outcomes: evidence from a recent English cohort. *Econ Educ Rev* 37:183-199

Ashiabi GS (2007) Play in the preschool classroom: its socioemotional significance and the teacher's role in play. *Early Child Educ J* 35:199-207

Baker M, Gruber J, Milligan K (2008) Universal child care, maternal labor supply, and family well-being. *J Political Econ* 116:709-745

Barnett WS (1995) Long-term effects of early childhood programs on cognitive and school outcomes. *Future Child* 5:25-50

Barnett WS, Jung K (2021) Effects of new jersey's abbot preschool program on children's achievement, grade retention, and special education through tenth grade. *Early Child Res Q* 56:248-259

Black SE, Devereux PJ, Salvanes KG (2005) The more the merrier? The effect of family size and birth order on children's education. *Q J Econ* 120:669-700

Bodrova E, Leong DJ (2015) Vygotskian and post-Vygotskian views on children's play. *Am J Play* 7:371-388

Boivin M, Vitaro F, Poulin F (2005) Peer relationships and the development of aggressive behavior in early childhood. In: Tremblay RE, Hartup WW, Archer J (eds) *Developmental origins of aggression*. The Guilford Press, pp 376-397

Bowles S, Gintis H, Osborne M (2001) The determinants of earnings: a behavioral approach. *J Econ Lit* 39:1137-1176

Cappelen A, List J, Samek A, Tungodden B (2020) The effect of early-childhood education on social preferences. *J Political Econ* 128:2739-2758

Caprara GV, Kanacri BPL, Gerbino M, Zuffiano A, Alessandri G, Vecchio G, Caprara E, Pastorelli C, Bridglall B (2014) Positive effects of promoting prosocial behavior in early adolescence: evidence from a school-based intervention. *Int J Behav Dev* 38:386-396

Currie J, Almond D (2011) Human capital development before age five. In: *Handbook of labor economics*. Elsevier, pp 1315-1486

Duckworth AL, Yeager DS (2015) Measurement matters: assessing personal qualities other than cognitive ability for educational purposes. *Educ Res* 44:237-251

Edin PA, Fredriksson P, Nybom M, Öckert B (2022) The rising return to non-cognitive skill. *Am Econ J Appl Econ* 14:78-100

Eivers AR, Brendgen M, Vitaro F, Borge AI (2012) Concurrent and longitudinal links between children's and their friends' antisocial and prosocial behavior in preschool. *Early Child Res Q* 27:137-146

Fabes RA, Carlo G, Kupanoff K, Laible D (1999) Early adolescence and prosocial/moral behavior I: the role of individual processes. *J Early Adolescence* 19:5-16

- Falbo T, Poston Jr DL (1993) The academic, personality, and physical outcomes of only children in China. *Child Dev* 64:18–35
- Gong X, Xu D, Han WJ (2016) The effects of preschool attendance on adolescent outcomes in rural China. *Early Child Res Q* 37:140–152
- Guo Y, Tang Y, Bai Y, Liu C (2024) Long-term benefits of early childhood education on off-farm employment: evidence from rural China. *Asia Pacific Education Review*, pp 1–23
- Gutman LM, Schoon I (2013) The impact of non-cognitive skills on outcomes for young people. *A Literature Review*
- Hampson RB (1984) Adolescent prosocial behavior: peer-group and situational factors associated with helping. *J Personal Soc Psychol* 46:153
- Havnes T, Mogstad M (2015) Is universal child care leveling the playing field? *J Public Econ* 127:100–114
- He L, Tan CS, Pung PW, Hu J, Tang HB, Cheng SM (2023) The role of parental rejection and poverty in the development of prosocial behavior among left-behind adolescents in rural China. *Child Youth Serv Rev* 155:107143
- Heckman JJ (2007) The economics, technology, and neuroscience of human capability formation. *Proc Natl Acad Sci USA* 104:13250–13255
- Heckman JJ, Stixrud J, Urzua S (2006) The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *J Labor Econ* 24:411–482
- Hirshberg D, Huang DSC, Fuller B (2005) Which low-income parents select child-care?: Family demand and neighborhood organizations. *Child Youth Serv Rev* 27:1119–1148
- Jencks C (1979) Who gets ahead? the determinants of economic success in america
- Ladd GW (2016) Themes and theories: perspectives on processes in family–peer relationships, in: *Family-peer relationships*. Routledge, pp 3–34
- Li H, Yao X, Zhang J, Zhou LA (2005) Parental childcare and children's educational attainment: evidence from China. *Appl Econ* 37:2067–2076
- Li L, Zhao L (2022) Does a “bad apple” spoil the bunch? The impact of low-achieving students on non-cognitive outcomes. *China Econ Rev* 76:101875
- Loeb S, Bridges M, Bassok D, Fuller B, Rumberger RW (2007) How much is too much? the influence of preschool centers on children's social and cognitive development. *Econ Educ Rev* 26:52–66
- Magnuson KA, Kelchen R, Duncan GJ, Schindler HS, Shager H, Yoshikawa H (2016) Do the effects of early childhood education programs differ by gender? A meta-analysis. *Early Child Res Q* 36:521–536
- Min W, Ru T, Shi Y (2019) The impact of early-life poverty experience on the non-cognitive abilities of rural adolescents: a life course perspective(in Chinese). *Contemporary Education Forum*
- Nagamachi M (2002) Kansei engineering as a powerful consumer-oriented technology for product development. *Appl Ergon* 33:289–294
- Nores M, Barnett WS (2010) Benefits of early childhood interventions across the world: (under) investing in the very young. *Econ Educ Rev* 29:271–282
- Rainford J (2020) Confidence and the effectiveness of creative methods in qualitative interviews with adults. *Int J Soc Res Methodol* 23:109–122
- Rauf FA, Tan WH (2020) Potensi realiti tirimuh dalam aktiviti mewarna: satu kajian di sebuah prasekolah: potential of augmented reality in colouring activity: a study in a preschool. *Southeast Asia Early Child J* 9:1–10
- Repper J, Carter T (2011) A review of the literature on peer support in mental health services. *J Ment Health* 20:392–411
- Reynolds AJ, Temple JA, Ou SR (2010) Preschool education, educational attainment, and crime prevention: Contributions of cognitive and non-cognitive skills. *Child Youth Serv Rev* 32:1054–1063
- Roberts BW (2009) Back to the future: personality and assessment and personality development. *J Res Personal* 43:137–145
- Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 39:33–38
- Schneider CR, Weber EU (2022) Motivating prosocial behavior by leveraging positive self-regard through values affirmation. *J Appl Soc Psychol* 52:106–114
- Tan M, Cai L, Bodovski K (2022) An active investment in cultural capital: structured extracurricular activities and educational success in China. *J Youth Stud* 25(8):1072–1087
- Tersi M, Matsouka O (2020) Improving social skills through structured playfulness program in preschool children. *Int J Instr* 13:259–274
- Tobin J, Hsueh Y, Karasawa M (2009) *Preschool in three cultures revisited: China, Japan, and the United States*. University of Chicago Press
- Wang D, Liu X, Xu G, Li K (2023) The impact of preschool experience on the long-term development of disadvantaged children(in Chinese). *Early Childhood Education Research*, pp 25–40
- Wang L, Zhang S, Rozelle S (2024) Do the pathways of child development before age three matter for development at primary school? Evidence from rural China[J]. *Human Soc Sci Commun* 11(1):1–10
- Wang T (2022) The peer effects of preschool: evidence from China's middle school. *Appl Econ* 54:1–18
- Wu KB (2011) Early childhood development and education in China: breaking the cycle of poverty and improving future competitiveness
- Xiao SX, Hanish LD, Malouf LM, Martin CL, Lecheile B, Goble P, Fabes RA, DeLay D, Bryce CI (2022) Preschoolers' interactions with other-gender peers promote prosocial behavior and reduce aggression: an examination of the buddy up intervention. *Early Child Res Q* 60:403–413
- Zhang G, Chayanuvata A et al. (2023) The impact of preschool education experience on multidimensional non-cognitive abilities and personality development of adolescents. *N Educ Rev* 71:101–113
- Zhang J, Zhou Y (2023) The peer effect of migrant children on students' non-cognitive outcomes: evidence from China. *Int J Educ Dev* 103:102883
- Zhang L (2013) Preschool experience, school readiness, self-regulation, and academic achievement: a longitudinal study in rural China. *HKU Theses Online (HKUTO)*
- Zhang L, Zhang L (2023) The impact of family structure on rural adolescents' delinquent behavior: Supervision vs. understanding—an empirical test based on the Chinese Education Panel Survey (CEPS) data (in Chinese). *Agricultural Technology and Economics*, pp 127–143
- Zhao C, Chen B (2022) Parental migration and non-cognitive abilities of left-behind children in rural China: causal effects by an instrumental variable approach. *Child Abus Negl* 123:105389
- Zhao L, Li L (2024) The impact of preschool education on children's cognitive and non-cognitive development: Evidence from China's preschool education popularization project (in Chinese). *Economics (Quarterly)*
- Zheng I, Weng Q, Gong X (2019) Preschool and the urban-rural cognition gap in middle school students: evidence from CEPs (in Chinese). *Sociological Studies*, pp 26
- Zheng L, Weng Q, Gong X (2021) Does preschool attendance affect the urban-rural cognition gap among middle school students? evidence from China education panel survey. *J Chin Sociol* 8:1–22

## Acknowledgements

The program was funded by the National Natural Science Foundation of China (Grant no. 72373085), Higher Education Discipline Innovation Project(Grant Number B16031), Humanities and Social Sciences Youth Foundation, Ministry of Education (Grant number 24JDSZK198), and the Shaanxi Provincial Social Science Foundation (Grant number 2023P023). The study's funder had no role in study design, data collection, data analysis, data interpretation, or report writing.

## Author contributions

W.L. curated the data and prepared the original draft. Y.L. interpreted the methodology and substantially revised the study, as well as developed the software. L.Z. performed data analysis and provided a partial interpretation of the results. Y.X. provided partial interpretations of the data analysis results. J.Z. contributed to the methodology and the interpretation of the data. H.G. conceptualized the study, developed the methodology, and was involved in writing and editing, as well as securing funding for the research. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Ethical approval

Research ethics approval for data collection in the China Educational Panel Survey (CEPS) dataset was granted by the Institutional Review Board of Renmin University, PR China. The survey was conducted in accordance with the ethical guidelines laid down in the Declaration of Helsinki.

## Informed consent

The consent was deemed not necessary. This study used the China Education Panel Survey (CEPS) data, and this data is fully open to the academic community. (<http://ceps.ruc.edu.cn/English/Home.htm>).

## Additional information

**Correspondence** and requests for materials should be addressed to Hongyu Guan.

**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.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025