# Pre-Analysis Plan for: AI-Assisted Parenting Guidance and Child Development: A Randomized Experiment in Rural China*

Pengfei Chu Xueheng Li Jia Wu Yiming Xia

February 13, 2026

*We present a pre-analysis plan for a cluster-randomized experiment that evaluates whether AI-assisted parenting guidance can improve parenting quality and child development during early adolescence. In a rural county in China, we randomly assign 78 primary-school classes (over 2,500 fourth- and fifth-grade children with parents at home from 10 schools) to one of three parental guidance arms: (i) a parenting guidebook adapted from UNICEF guidelines, (ii) an AI-assisted parenting chatbot built on a trained large language model that delivers real-time personalized advice, or (iii) a no-guidance control. In a cross-cutting design, parents within each class are further randomized to receive SMS messages conveying children's expressed expectation of their parents, the benefits of supportive parenting, or a perspective-taking reflection message. All information was sent to high- and low-message coverage classes. We examine the treatment effects on parenting styles and practices and on children's mental health and academic achievement. This study provides among the first experimental evidence on parenting interventions and on whether AI-enabled tools can enhance parenting quality and children's outcomes.* (*JEL* I21, I28, J13)

*Keywords*: Parenting style and practice, AI-assisted parenting, Child development, Mental health, Academic achievement

*Chu: South China Normal University (e-mail: pfchu@outlook.com); Li: Sun Yat-sen University (e-mail: xuehenglee88@gmail.com); Wu: Sun Yat-sen University (e-mail: wujia@mail.sysu.edu.cn); Xia: Southwestern University of Finance and Economics (e-mail: xiaym@swufe.edu.cn). The trial received IRB approval from the Research Ethics Committee of the China Center for Behavioral Economics and Finance (CCBEF), approval ID: 2025B05.

**Timeline**: The baseline survey and test scores were collected in September 8, 2025. The interventions began on September 15, 2025, and ended on January 31, 2026. The follow-up survey took place in February 2026, and the follow-up test scores will be gathered in early March 2026. After data entry and cleaning, the authors are expected to have access to endline data from May 2026. Our expected date for completion of the pre-specified research design and disposition of the Stage 2 report is January 2027.

# 1 Introduction

## 1.1 Motivation

Parental engagement is a central determinant of child development, with evidence from both economics and developmental psychology showing consistent associations with improvements in cognition, socioemotional skills, and educational achievement (e.g., Carneiro and Heckman, 2003; Fan and Chen, 2001). Parenting programs have been especially effective in the early years: landmark interventions, such as Jamaica's home-visiting program, Perry Preschool, and Abecedarian, produced lasting gains in cognition, schooling, and earnings (Gertler et al., 2014; Heckman et al., 2010).

By contrast, evidence on effective parenting during early adolescence (ages 10–13 according to the American Academy of Pediatrics) remains limited, particularly in low- and middle-income countries (LMICs). This period coincides with rising mental health challenges and the emergence of emotional and behavioral problems (Almond et al., 2018; Ward et al., 2020; Currie, 2025). Among factors contributing to these outcomes, parenting plays a role and also could act as a counterforce to alleviate these problems (Doepke and Zilibotti, 2017; Edlund, 2021), yet many parents fail to recognize this. In particular, many parents lack the skills to communicate with and guide their children effectively during this stage, potentially exacerbating existing mental health difficulties (Steinberg, 2001).

The challenges associated with low-quality parenting are particularly relevant in rural China. These parents usually show a high demand for their children's academic performance. However, possibly due to their own limited education or childhood experiences in which obedience was emphasized, they are not listening to what their children say. Consequently, when children exhibit misbehavior, some parents resort to scolding and punishment.[1] Within Baumrind's framework, such an imbalance between demand and responsive-

[1]Such parenting behaviors, characterized by high demands paired with low emotional responsiveness, are classified as authoritarian parenting (Maccoby and Martin, 1983), which is well illustrated by the "Tiger Parenting" (Chua, 2011).

ness is a well-established predictor of psychological distress (Baumrind, 2013).

In addition, an emphasis on strict obedience may hinder the development of other important skills, such as creativity and prosociality, which are increasingly valued in modern societies. Prior research links these negative parenting environments to weaker parent–child relationships, poorer socioemotional development, and higher internalizing symptoms during adolescence (e.g. Steinberg, 2001; Repetti et al., 2002).

This context highlights the need for scalable, culturally appropriate interventions that help parents shift toward more supportive practices. Such interventions should maintain the educational performance while fostering open communication and creating space for children to express their views and develop autonomy and creativity.

A conventional, low-cost option is to distribute booklets that translate evidence-based advice from psychologists and educational experts into practical guidance for parents. While this approach is effective at providing general principles, these static materials may fall short in addressing the specific situations parents face in daily interactions with children. Recent advances in AI technology offer a promising solution. AI chatbots, after being trained by learning evidence-based parenting principles, can deliver real-time, context-specific, and personalized guidance to help parents communicate more effectively and resolve conflicts, thereby strengthening parenting practices and improving child outcomes (e.g. Alam et al., 2023). Despite this promise, empirical evidence on AI-assisted parenting remains scarce.

In collaboration with the Local Education Bureau, we recruited ten schools in a County in Hunan Province to participate in our experiment. Our baseline sample consists of over 3,500 fourth- and fifth-grade students, including 2,505 non-left-behind children from 78 classes in 10 public primary schools. Our interventions target parents of these 2,552 non-left-behind children. We also analyze spillover effects on left-behind children.

We use a two-level cross-randomization design. First, within each school-by-grade stratum, we randomly assigned classes to two parental guidance treatment groups or the control group. Second, within each class, we randomly assigned parents to one of three informa-

tion nudge treatments or to a control. This design allows us to estimate the effects of each parental guidance treatment, each information treatment, and their interactions.

We design two parental guidance interventions: (i) *parenting guidebook (PG)*. Parents received a guidebook on how to tackle parenting challenges. This guidebook, adapted from UNICEF's parenting education guidelines and tailored to the Chinese cultural context, includes practical examples related to ten key issues identified by local parents in a preliminary survey. (ii) *AI-assisted parenting chatbot (AIPC)*. Parents were provided with an account to access our designed app, which is based on a trained large language model (Deepseek). This AI chatbot allows parents to acquire real-time, personalized advice for daily parenting challenges. The chatbot was trained using the same material as the PG intervention, enabling it to offer suggestions in line with UNICEF's guidelines. We assigned 26 classes to PG treatment groups and 26 classes to AIPC treatment groups. The remaining 26 classes were assigned as control classes (between-class controls) with no parenting guidance.

We then cross-randomized students across the three parental guidance intervention arms and the information treatment arm. Each student (and their parents) was randomly assigned to one of three information treatment groups or to a control group: (i) *Children's expressed needs*: Parents received information about children's expressed preferences for parental engagement, derived from the baseline survey. (ii) *Benefits of supportive parenting*: Parents received information on the positive effects of patient and egalitarian parenting practices. (iii) *Perspective-taking reflection*: Parents received a message encouraging recalling their own negative childhood experiences like corporal punishment. (iv) *Control*: Parents in the control group received no information messages. We refer to this group as *within-class controls*. All information was delivered to parents twice a month via SMS messages.

Moreover, to examine the possible spillover effects of information delivery, we design high- and low-information coverage classes. Specifically, within each PG, AIPC, and between-class controls, we randomly assigned half of the classes to receive high-coverage information and the other half to receive low-coverage information. In the high-coverage

information classes, children in a class were equally distributed across four information treatment groups with equal probability: 25% of children's parents received the children's expressed needs message, 25% received the benefits of supportive parenting message, 25% received the perspective-taking reflection message, and 25% served as individual controls. In the low-coverage information classes, the distribution was unequal: 13% of children's parents were assigned to receive the children's expressed needs message, the benefits of supportive parenting message, and the perspective-taking reflection message, while 61% of parents served as controls.

The cluster-randomized design allows us to estimate intention-to-treat (ITT) effects. Our primary analysis examines how the parental guidance intervention, the information intervention, and their interaction affect parenting styles and practices, which are expected to further influence children's mental health and academic achievement. In secondary analyses, we assess whether AI-assisted guidance outperforms traditional booklet guidance, thereby testing whether real-time and personalized advice provide additional benefits in parenting. We also examine intermediate outcomes, including AI usage, the parent-child relationship, parenting knowledge, and parental educational investment, to test the program's Theory of Change (see Appendix Section A) and to identify potential mechanisms or points of breakdown if the intervention fails to generate the expected impacts.

Our contribution to the literature is twofold. First, we study parenting programs that target child development during the "missing middle" years (Almond et al., 2018), a critical period that has received far less attention than early childhood but is increasingly salient given the rise of school-age adolescent mental health problems (e.g., anxiety, depressive symptoms, and emotional/behavioral difficulties) that schools and families struggle to manage (Almond et al., 2018; Ward et al., 2020; Currie, 2025). While extensive evidence shows that parenting interventions can promote early childhood development across diverse contexts and modalities (e.g., York et al., 2019; Justino et al., 2023; Carneiro et al., 2024; Attanasio et al., 2024), only a handful of recent studies examine how parents can be engaged during

adolescence, particularly through low-cost smartphone-based interventions (e.g., Cunha et al., 2023). We address this gap by documenting the challenges parents face and by testing whether traditional or AI-assisted parenting tools can strengthen parenting skills, thereby providing rare evidence on how such programs can support families and foster adolescent development.

Second, our study contributes to the growing literature on AI and education, where rigorous evidence remains limited and mixed. Most existing research examines AI tools designed to assist students directly. Some studies report positive impacts of adaptive learning technologies, suggesting that AI can tailor instruction to individual needs and improve academic outcomes (e.g., Muralidharan et al., 2019; De Simone et al., 2025). Other evaluations of AI tutors, however, find null or even negative effects on certain dimensions of learning, such as critical thinking (e.g., Bastani et al., 2024). Much less is known about whether AI applications can empower and inform parents—an area with considerable promise for strengthening human capital development.

## 1.2 Research questions

We study whether providing parenting guidance, delivered either through a parenting guidebook (PG) or an AI-assisted parenting chatbot (AIPC) and, in some arms, complemented by SMS information messages, can improve parenting practices and children's academic achievement and mental health. Specifically, the study addresses the following primary questions:

(i) Do PG and AIPC treatments, or complemented by SMS information messages improve parenting styles and practices?

(ii) Do improvements in parenting styles and practices translate into gains in children's academic performance and mental health?

We also investigate the following secondary research questions:

(i) What is the take-up of each intervention?

(ii) Does AI-assisted parenting outperform the traditional booklet-based intervention?

(iii) Do our treatments affect parent-child relationships, children's socio-emotional skills, other non-cognitive skills (e.g., "Big-five" personality traits, misbehavior, growth mindset), creativity, and friendship networks?

(iv) Do our treatments affect parental time inputs and educational expenditure?

(v) Do these parenting interventions generate spillover effects for left-behind children?

# 2 Data Collection

## 2.1 Background

Our experiment was conducted in a rural county in Hunan Province. In 2024, the county's per capita GDP was 35,769 RMB, which was only 37% of the national average for the same period. Enrollment in primary school is nearly universal, thanks to the Compulsory Education Law of China. However, parents' education levels remain limited: parents in our sample completed about 11 years of schooling on average, only two years more than the nine years mandated by compulsory education.

Relatedly, many parents lack modern parenting knowledge and still rely on traditional approaches that emphasize obedience. Many parents do not have the patience to listen to their children. Deeply rooted traditional Confucian beliefs that emphasize parental authority exacerbate this pattern, leading some parents to resort to corporal punishment. In 2024, UNICEF (2024) reported that six in ten children under 5 years of age regularly suffer physical punishment and/or psychological violence at the hands of parents and caregivers. According to our baseline survey, parents reported that they experienced conflicts between their children nearly every week. Children reported being physically punished or verbally abused by their parents once every two to three weeks.

## 2.2 Sample and survey

### 2.2.1 Analyzed sample

The baseline survey includes 3,552 fourth- and fifth-grade students from 10 primary schools. We used the baseline survey to determine parental absence status. Of these students, 2,505 have at least one parent at home (non-left-behind children). Since our study focuses on the impact of improved parenting style, both the parenting guidance and information nudge interventions targeted these 2,505 non-left-behind children. When analyzing spillover effects of the treatments, we also consider outcomes for left-behind children. The baseline survey asked about each student's main caretaker and their contact information so we could send them SMS messages, a guidebook, or login information of our AI chatbot, depending on their treatment status.[2]

### 2.2.2 Survey

We implemented two waves of surveys. The baseline questionnaire, conducted in September 2025, was designed to measure parent and child characteristics and elicit parenting styles and practices, key outcome variables of children, and other variables of interest (e.g., first-stage outcomes of parents' usage of our designed AI chatbot, educational investment, and children's non-cognitive skills and creativity). The follow-up survey will be conducted in February 2026. It will include all questions from the baseline survey. We expand the questionnaire in the follow-up survey by adding parents' assessments of our designed booklet and AI-assisted chatbot. We also include a short five-question test to measure parents' knowledge of the booklet content.

We designed questionnaires for students and parents, which all students, including left-behind children, were required to complete. The completion rate was high. Fewer than 20 children with severe intellectual disabilities did not finish the survey. We collected question-

[2]For non-left-behind children, their main caretaker is the parent.

naires from 3,552 students and their parents.

*Student questionnaire*.—We collected demographic characteristics and baseline measures of (i) mental health using the CES-D-10 scale, (ii) academic performance measured by average Chinese and mathematics test scores obtained from the Education Bureau, (iii) other outcomes, including misbehaviors, socio-emotional skills, Big Five personality traits, (iv) creativity, and (v) time diary recording time spending on various activities at home. In addition, we measure children's perceptions of parenting practices and their expectations of parenting (e.g., whether they hope parents listen to them, hope to express their views) and children's reports of parent–child conflict, including verbal abuse and physical punishment.

*Parent questionnaire.*—We collected (i) parents' demographic and socioeconomic characteristics (including parental education, employment, income, and household expenditures), (ii) parenting style and practices (including parents' self-reported parenting styles, discipline practices, positive parenting, number of parent–child conflicts measured in 10 situations, parents' perceived efficacy of various parenting practices, parents' recollections of the parenting practices they experienced during childhood), (iii) parental expenditure and time investment in children's education, (iv) parents' mental health (CES-D-10), and (v) AI adoption for parenting and their perceived roles of AI tools. (vi) Parents' workplace location, which is used to identify non-left-behind children.

Primary data collection is conducted by trained university students who are external to the study schools, with logistical support from the schools. To ensure quality, we follow strict protocols that include spot-checks and accompanied visits.

Students were asked to complete their section at school, while the questionnaires of the parents' section were collected the next day when the parents finished filling out the questionnaire. After students and parents complete the questionnaires, our research team oversees and verifies that the answer sheets are filled in correctly. The completed sheets are then digitized using a scanner.

## 2.3 Measure of outcomes

### 2.3.1 Main outcomes

We focus on two main outcomes: children's *academic performance* and *mental health*.

*Academic performance*.—Children's academic performance is measured by Chinese language and mathematics test scores. At the end of each school semester, each school administers an exam to assess students' academic performance. We obtain transcript data for each end-of-semester exam before and after the interventions from the Educational Bureau (The second-round test scores will be collected in early March 2026). We standardize the raw test scores in each school-grade cell according to the between-class control group to have a mean of 0 and a standard deviation of 1 by wave. As complementary measures, we also use parents' assessments of children's academic performance and parents' grade expectations.

*Mental health*.—Children's mental health is measured by the CES-D-10 scale. Students report how often, during the past week, they experienced each of ten feelings or behaviors (e.g., depressed mood, difficulty concentrating, poor sleep, loneliness). Each item is coded on a four-point frequency scale (0–3), with two positively worded items reverse-coded so that higher values consistently indicate worse mental health. We sum the ten items to obtain a total score ranging from 0 to 30. A score of 10 or above indicates an elevated risk of depression.

### 2.3.2 Intermediate outcomes

We measure four categories of intermediate outcomes that capture the mechanisms through which the intervention may affect children's academic performance and mental health.

*Parenting style and practice.*—Parenting style and practice are assessed from both child and parent perspectives. Children report on parenting style using items adapted from Maccoby and Martin (1983), rating each parent's behaviors (e.g., encouraging expression of

thoughts, explaining rules, offering comfort, using physical punishment, scolding) on a five-point Likert scale. Parents also reported on discipline strategies, rule enforcement, positive parenting behaviors, perceived efficacy of different parenting approaches, and parental empathy.

*Parent–child relationship quality.*—The parent–child relationship is assessed by the frequency of conflicts reported by parents and by children's reports of verbal abuse and physical punishment.

*Parents' AI chatbot and guidebook usage.*—AI chatbot is measured through app-recorded chatbot usage logs (available for the AI treatment group). Guidebook usage is proxied by a five-question knowledge test assessing parents' comprehension of the guidebook content.

*Parental educational inputs.*—Parental time inputs are measured by children's reports of time spent on various activities, including accompanying children's study, reading stories to children, playing with children, etc. All these metrics are reported in minutes per day. Parental expenditure is measured by parents' reports of monthly educational expenditure across categories (e.g., extracurricular tutoring, books, interest classes).

### 2.3.3 Other outcomes

We also measure four categories of other outcomes: (i) socio-emotional skills, captured by SDQ domains (Goodman, 1997) ; (ii) non-cognitive skills, including economic preference ( delay of gratification, risk preference, and altruism), Big Five personality traits (BFI-10) (Rammstedt and John, 2007), self-control, misbehavior, and growth mindset (Dweck, 2006); (iii) self-perceived creativity; and (iv) peer networks, measured by number of friends and perceived peer relationship quality.

To create outcome variables for evaluation, we aggregate items within each family and standardize scores relative to the control group so that they have a mean of 0 and a standard

deviation of 1. Table 1 summarizes the outcome variables. Appendix Section C provides detailed information about the survey items used to construct the above outcome variables.

# 3 Experiment Design

## 3.1 Randomization

Ten rural primary schools are included, with a sample of 3,552 fourth- and fifth-grade students (2,505 non-left-behind children) across 78 classes and their parents. Our interventions target the non-left-behind children's parents (main caretaker) who are at home during the experiment. The ten schools were not randomly selected. Schools were eligible if they had at least three classes in grades 4 and 5, which allows class-level assignment within school-by-grade strata in the first stage of randomization. We implement a two-level cross-randomized cluster randomized controlled trial.

In the first stage, we design parental guidance interventions. Classes were randomly assigned to one of three parental-guidance arms: parenting guidebook (PG), AI-assisted parenting chatbot (AIPC), or no-guidance control (between-class comparison). The 78 classes were stratified by school and grade, then split across the three arms with equal probability. Children's main caretaker received either the parenting guidebook or the login information to access our trained AI chatbot.

In the second stage, we design information nudge interventions. Information is delivered to parents of non-left-behind children via SMS. Conditional on the treatment arm assigned in the first stage, classes are first randomly assigned to high-coverage or low-coverage groups based on the intended intensity of message exposure. Next, within each PG, AIPC, and between-class control arm, we randomly assigned half of the classes to receive high-coverage information and the other half to receive low-coverage information. In high-coverage classes, parents were equally distributed across four information treatment groups: 25% received the children's expressed needs message, 25% received the benefits of

supportive parenting message, 25% received the perspective-taking reflection message, and 25% served as within-class controls. In low-coverage classes, the distribution was unequal: 13% of parents were assigned to receive each of the three message types, while 61% of parents served as within-class controls.

## 3.2 Intervention

### 3.2.1 Parental guidance intervention

We design two parental guidance interventions and a between-class control. The parental guidance intervention began on September 15 2025, and ended on January 31 2026.

- *Treatment 1: Parenting guidebook (PG).*— Parents received a guidebook on how to tackle parenting challenges. This guidebook, adapted from UNICEF's parenting education guidelines and tailored to the Chinese cultural context, includes practical examples related to ten key issues identified by local parents in a preliminary survey. Starting in Week 2, parents also received biweekly reminders encouraging continued use of the materials. Figure 1 displays the guidebook.

- *Treatment 2: AI-assisted parenting chatbot (AIPC).* Parents were provided with an account to access our designed app, which is based on a trained large language model (Deepseek). This AI chatbot allows parents to acquire real-time, personalized advice for daily parenting challenges. The chatbot was trained using the same material as the PG intervention, enabling it to offer suggestions in line with UNICEF's guidelines. We pre-registered accounts for treated parents to ensure access and distributed a user manual that explains how to interact with the tool. Parents were first coached on how to use the tool through a live demo. They then downloaded the app to their phones using a pre-registered account uniquely assigned to them, ensuring exclusive access. From Week 2, parents received reminders twice a month encouraging use of the AI chatbot. Figure 2 displays the login page of the AI-assisted chatbot app.

- *Control group (between-class control).* No intervention.

### 3.2.2 Information nudge intervention

We design three information nudge interventions and a within-class control. Randomly selected parents receive the messages that were delivered twice a month from 1 October 2025 to 31 January 2026.

- *Children's expressed needs*. [XXX County Education Bureau] Dear Parents, in September, we surveyed students at your child's school. The results show that [XX]% of students hope their parents will show more understanding and respect, [YY]% want more space to share their own ideas, and [ZZ]% would like more opportunities to communicate with their parents.[3]

- *Benefits of supportive parenting*. [XXX County Education Bureau] Dear Parents, in September, we surveyed students at your child's school. The results suggest that treating children with patience and respect can reduce anxiety by [X]% and strengthen independent thinking by [Y]%.[4]

- *Perspective-taking reflection*. [XXX County Education Bureau] Dear Parents, many of us remember growing up with harsh discipline and being expected to obey without question—and how unpleasant that felt. We encourage you to see things from your child's perspective: offer more understanding and respect in daily interactions, and make time to talk and listen.

- *Within-class control*. No information.

To alleviate parental concerns regarding the guidebook and chatbot's credibility, both were officially endorsed by the local education bureau, bearing its name and logo.

[3] [XX]% varies across schools with a mean of 39% (range: 33–45%). [YY]% has a mean of 70% (range: 58–78%). [ZZ]% has a mean of 72% (range: 62–80%).

[4] [X]% has a mean of 13.61% (range: 5.6–19.1%). [Y]% has a mean of 12.43% (range: 6.1–20%).

We designed the information nudge intervention to be cross-randomized with the parental guidance treatment for three reasons. First, parents in the PG and AIPC groups may use guidance tools to address their children's issues appropriately if they underestimate the impact of parenting on child development or recognize their children's needs for more parent-child communication. Information nudges that highlight children's expectations, demonstrate the benefits of supportive parenting practices, or encourage perspective-taking can help parents recognize the value of positive parenting. This may motivate them to improve their skills directly or refer to guidance tools for support. Second, cross-randomizing the information and guidance treatments enables us to estimate not only the main effects of each intervention but also their interaction effects, allowing us to test whether information provision complements or substitutes for parenting guidance. Third, we vary the intensity of information coverage at the class level using high and low coverage rates to examine potential spillover effects on control students (within-class control) and left-behind children. If information treatments produce spillover effects through peer interactions or changes in the classroom environment, we expect these effects to be greater in classes with higher coverage.

In sum, this relatively complex design enables us to evaluate the effectiveness of delivery methods and the impact of information/reminders on promoting adoption and sustained use of guidance tools, so that we can address challenges in real-world parenting interventions.

Figure 3 illustrates our experiment design.

### 3.2.3 Implementation

During the experiment, enumerators and administrative data providers are masked, which ensures the validity of survey data collection and strengthens the credibility of administrative outcomes as unbiased measures.

Parents and students know only their own assignment (by receiving either the booklet or the AI chatbot) and are not told that other arms exist. Some parents or students in the

control group may realize that they did not receive any parenting materials. However, this is unlikely, as class head teachers regularly send homework tasks to parents, and such variation in school–home communication is considered normal.

## 3.3 Attrition and spillover

*Attrition.*—Previous studies in the same county show that attrition is mainly due to school transfers and typically does not exceed 9 percent. We will assess whether our intervention affects attrition rates. Additionally, attrition has already been accounted for in our statistical power calculations.

*Spillover.*—We believe spillovers between experimental groups are unlikely. For the parenting guidance interventions (PG and AIPC), randomization was done at the class level, and the intervention targeted parents directly. In China, primary school students stay in their assigned classes without cross-class learning, and parents from different classes rarely interact. Additionally, access to the AI tool is limited to parents in the AI group by providing each child-parent pair with a single account, ensuring that parents in the Control or Booklet groups cannot use it. These measures help minimize the risk of spillover across classes. To further assess potential spillovers of the parenting guidance interventions, we can exploit variation in classroom proximity by examining whether classrooms are located on the same floor, allowing us to test whether physical proximity affects treatment diffusion.

For the information nudge treatment, we examine potential spillovers by design through the randomized variation in coverage rates within classrooms. By comparing outcomes across high- and low-coverage classes, we can directly estimate within-class spillover effects of information messages on untreated parents in the same classroom.

### 3.4 Challenges in data collection

We expect only minor attrition (less than 10%), based on experience from previous studies, as well as the possibility of some outliers. We will document attrition, test for its selectiveness, and address it using standard methods. Outliers will be handled with rank-based measures, as described in the statistical methods section.

Non-compliance with the assigned treatment is likely. However, because the intervention is relatively simple and beneficial for parenting, we expect a higher-than-usual take-up rate. Additionally, our information nudge may further increase participation. Our primary analysis uses the intention-to-treat (ITT) estimator, which remains valid even with non-compliance and captures the causal effect of being assigned to treatment. Since we can track parental use of the AI chatbot through usage logs on our online platform, we are also able to examine the average treatment effect on the treated (ATT).

## 4 Findings from baseline

### 4.1 Summary statistics from the student survey

We begin by plotting the baseline distributions of our primary outcomes by treatment arm. Panel A of Figure 4 plots the kernel distribution of mental health for the PG, AIPC, and Control groups. Mental health is measured using the CES-D 10 scale, where higher scores indicate more severe depressive symptoms. The dashed line represents the cutoff score of 10, indicating depressive symptoms and identifying individuals at risk. We find a substantial proportion of children had a score higher than 10. Moreover, the figure shows that the distributions are similar across the three treatment and control groups, confirming that randomization was balanced. Panels B and C in Figure 4 display kernel densities of standardized Chinese and math test scores for treatment and control groups. As with the mental health measure, the three groups track each other closely.

Table 2 reports the summary statistics for variables constructed from the students' survey. We focus on 2,505 children with at least one parent at home. Panel A shows children's demographic characteristics and baseline health conditions. The sample is gender-balanced, with 53% boys, and children are on average 10 years old. Most children have a rural *hukou* (72.4%) and are non-minority (97.8%). The average number of siblings is 1.46. Children report an average of 1.5 illness episodes in the last semester.

Panel B reports children's outcomes. Mean Chinese and math test scores are approximately 78. Mental health, measured as CES-D 10, has a mean of 8.7, implying that children in our sample border on relevant depressive symptoms with the cutoff of 10. This suggests that mental health issues are an important concern among the children in our sample.

Children report an average of 7.8 instances of verbal abuse and physical punishment in the last semester. This means that, on average, children experienced verbal abuse or physical punishment approximately once every 2-3 weeks during the semester. The high maximum value of 28 suggests that some children faced such mistreatment more than once a week. Our results indicate that a subset of children is exposed to frequent verbal and physical mistreatment, which may negatively affect their psychological well-being and development

On a 1-to-4 scale, children's need to express themselves to parents averages 2.93, and their need for parental listening averages 2.92. Perceived parental misunderstanding averages 2.16. Together, these patterns are consistent with the fact that children experienced frequent conflict with parents and thus express a strong desire for better communication. It also suggests that many parents may lack effective strategies for engaging with their children, reinforcing the motivation for a parenting intervention.

The table also documents variation in children's other developmental outcomes. Self-control averages 11.7 out of 20, and prosocial behavior averages 4.7 out of 6. Among the Big Five personality traits, agreeableness scores the highest and conscientiousness the lowest. Self-perceived creativity averages 26.5 out of 36. Growth mindset averages 11.6 out of 15. Children report an average of 30.9 misbehaviors in the last semester and nominate an

average of 5.2 friends.

Panel C reports parenting styles as perceived by children, classified following Maccoby and Martin (1983) into authoritative, authoritarian, permissive, and neglectful categories. Each style is measured by a continuous index. For both fathers and mothers, authoritative parenting scores the highest (10.5 for fathers, 11.1 for mothers), followed by permissive, authoritarian, and neglectful. The ranking is consistent across parents, though mothers receive slightly higher authoritative scores than fathers.

How do parenting styles relate to children's mental health and academic performance, especially given that their average CES-D scores are near the clinical distress threshold? Table 4 shows that authoritative parenting is linked to lower psychological distress and better academic outcomes. Authoritarian parenting is associated with higher test scores but also greater psychological distress. Maternal permissive and neglectful parenting styles correlate with significantly lower academic performance. Additionally, maternal parenting styles have stronger and more consistent associations with test scores than paternal styles. These patterns motivate the focus of our intervention: encouraging parents to provide greater emotional support, shift from authoritarian toward authoritative practices, and thereby reduce children's psychological distress.

Panel D of Table 2 reports children's daily time allocation. Children spend the most time on homework, reviewing, and previewing lessons (45.6 minutes per day), followed by playing with friends (37.9 minutes), studying with parental supervision (31.5 minutes), and doing sports (29.6 minutes). By contrast, children spend only 17.7 minutes playing with parents and 20.3 minutes listening to stories told by parents, suggesting limited direct parent-child engagement outside of academic supervision.

We present variables derived from the children's survey for all 3,552 children (including left-behind children) in Appendix Table D1.

## 4.2 Summary statistics from the parent survey

Table 3 presents summary statistics from the parent survey. Panel A shows that fathers average about 40 years old and mothers about 37, with both having completed roughly 11.5 years of schooling, consistent with a moderately educated rural population. Monthly household income averages approximately 6,900 CNY, with a large dispersion that reflects substantial economic heterogeneity.

Panel B shows parents' AI usage and their views on its role in parenting. AI adoption for parenting is limited: only 6.3% of parents use AI tools for their children's education, compared to 27% who use them for work or study. This indicates that AI tools are not yet widely used for tutoring children. The average score for parents' belief in the usefulness of AI in parenting is 19 out of 28.

Panel C presents parents' evaluation of their children. On average, parents reported a satisfaction level of 7.11 (out of 10; SD = 1.868) with their child's academic performance in the last semester. Parents rated their child's academic performance at 76.42 (out of 100; SD = 19.056). The average expectation for the child's academic achievement was 2.68 (out of 3; SD = 0.502), suggesting that while parents were generally satisfied with children's academic performance, they still had higher expectations. Parents' mental health, measured by the CES-D 10, averages 8.15 (SD = 4.25), similar to children's reported average of 8.7.

Panel D presents parents' outcomes and their perceptions of their own parenting practices. On average, parents reported 18 parent–child conflicts (SD = 14.25) in the past semester. They indicated a relatively high level of 8.15 (out of 10; SD = 1.41) for rule enforcement, consistent with Chinese cultural traditions that emphasize obedience. Parents also reported an average emotional regulation score of 7.44 (out of 10; SD = 1.38) and an average communication skills score of 7.88 (out of 10; SD = 1.65).

Positive parenting practices averaged 15.53 (out of 20; SD = 2.954), while parenting self-efficacy scored 21.75 (out of 30; SD = 4.104), and parenting awareness averaged 30.946 (out of 40; SD = 7.471). Parents demonstrated an empathy level of 14.4 (out of 20; SD =

2.867).

Regarding discipline measures, parents reported that communicative discipline was most frequently used (mean = 6.45; SD = 2.625), followed by command-style (mean = 1.256; SD = 1.674) and punitive discipline (mean = 0.998; SD = 1.534), while permissive discipline was least common (mean = 0.319; SD = 0.877). Yet as we have seen in Table 2, children's reports reveal a sizable gap between how parents evaluate their own communication and children's experience. Despite parents' strong self-evaluation on communicative discipline, children still report high levels of unmet need for self-expression and parental listening, along with notable perceived parental misunderstanding.

Panel E shows households spending on children's education and leisure, with monthly expenses for school supplies, extracurricular resources, and toys averaging 54–76 CNY each, approximately 1% of household income. Panel F reports grandparents' parenting styles during the current parents' own childhoods. Both paternal and maternal grandparents had high scores on authoritative and authoritarian parenting.

We present variables constructed from the parents' survey for all 3,552 children (including left-behind children) in Appendix Table D2.

## 4.3 Balance test

Columns (1)–(2) of Appendix Tables 5 and 6 show that baseline characteristics and outcomes are well balanced across the three main treatment arms. We find no systematic differences in children's demographics, baseline outcomes, parenting styles, or parental time investments across treatment groups, indicating that the randomization successfully generated comparable groups at baseline. Specifically, for variables constructed from the child survey, Table 5 shows that only 1 out of 48 variables is statistically significant at the 5% level when comparing the AIPC group with the control (Column 1), and 3 out of 48 are significant at the 10% level when comparing the PG group with the control (Column 2). For variables from the parent survey, Table 6 shows that 0 (1) out of 34 variables is significant

at the 10% level for the AIPC-control (PG-control) comparison. These results support that randomization achieved balance across treatment arms.

In addition, columns (3)–(6) of Appendix Tables 5 and 6 show that balance is still maintained across three information nudge treatment groups and within-class control groups.

# 5 Empirical analysis

## 5.1 Statistical method

We use intention-to-treat (ITT) as the primary analysis, which does not require full compliance and is interpreted as the effect of assignment to treatment. To complement this, we also estimate ATT effects using two-stage least squares (2SLS).

### 5.1.1 Effect of Main Treatment Arms

*Effect on parenting style and practice.*—Among intermediate outcomes listed in Table 1, parenting styles and practices are of primary interest, as they capture the parents' behaviors most directly targeted by both interventions. We first test whether the interventions work for improving parenting practices and style. We estimate the ITT effect on each parenting practice and style measure $j$ using the following specification:

$$PS^j_{icgs} = \beta^j_1 T^{PG}_c + \beta^j_2 T^{AIPC}_c + \delta^j X_{icgs0} + \xi_{gs} + \epsilon^j_{icgs}, \quad j = 1, \ldots, J \tag{1}$$

where the subscripts $i$, $c$, $g$, and $s$ denote child, class, grade, and school, respectively. $PS^j_{icgs}$ is the standardized score for parenting practice or style measure $j$ of child $i$ in class $c$ grade $g$ at school $s$. The index $j$ runs over variables related to parenting style and practice from Table 1. Specifically, from the child survey, we measure parenting style and the frequency of conflicts with parents; from the parent survey, we measure discipline strategies, rule enforcement, parent empathy, positive parenting, perceived efficacy of parenting practices, and

their reported frequency of conflicts with children. We also examine parental educational inputs. $T_c^{PG}$ and $T_c^{AIPC}$ denote class-level assignment to the parenting guidebook and AI-assisted parenting chatbot arms, respectively; $X_{icgs0}$ includes baseline covariates and the lagged dependent variable; $\xi_{gs}$ are strata (school-by-grade) fixed effects. The coefficients $\beta_1^j$ and $\beta_2^j$ capture the ITT effects of the two interventions on parenting outcome $j$. Standard errors are clustered at the class (randomization) level.

*Children's outcome.*—Next, we examine whether changes in parenting styles and practices lead to improvements in children's outcomes. We use the following model:

$$\textit{Child outcome}_{icgs}^j = \gamma_1^j T_c^{PG} + \gamma_2^j T_c^{AIPC} + \delta^j X_{icgs0} + \xi_{gs} + \epsilon_{icgs}^j, \quad j = 1, \ldots, J \tag{2}$$

where $\textit{Child outcome}_{icgs}^j$ denotes any outcome of child $i$ or of the child's parents, including main outcomes (children's mental health and academic performance), and additional outcomes (socio-emotional skills, non-cognitive skills, creativity, peer networks). All other notation follows from above.

### 5.1.2 Local Average Treatment Effect (LATE)

For the AI-assisted parenting chatbot (AIPC) arm, compliance is imperfect. Not all parents assigned to the chatbot actually use it. Because the chatbot is delivered through our designed app, we observe each parent's actual usage, including whether and how frequently they interacted with the chatbot, from administrative log data. This allows us to estimate the local average treatment effect (LATE) using two-stage least squares (2SLS), with random assignment to the AIPC arm as an instrument for actual usage. The first-stage equation is:

$$\textit{AI usage}_{icgs} = \pi T_c^{AIPC} + \gamma' X_{icgs0} + \xi_{gs} + u_{icgs}, \tag{3}$$

and the second-stage equation is:

$$Y_{icgs} = \pi^{LATE} \widehat{\textit{AI usage}}_{icgs} + \delta' X_{icgs0} + \xi_{gs} + \eta_{icgs}. \tag{4}$$

Here, $Y_{icgs}$ can be $PS^{j}_{icgs}$ or $\textit{Child outcome}^{j}_{icgs}$. $\textit{AI usage}_{icgs}$ denotes actual AI chatbot usage of child $i's$ parent, measured either as a binary indicator or as a continuous usage frequency variable, both constructed from administrative logs on our website; $\textit{AI usage}_{icgs} = 0$ for all individuals in the control group. The coefficient $\pi^{LATE}$ identifies the LATE of AI chatbot usage for compliers, namely, parents who use the chatbot when assigned to the AIPC arm but would not use it otherwise.

### 5.1.3 Effect of Information Provision

To estimate the treatment effects of information nudge with different coverage rates and message contents while accounting for the main experimental arms, we estimate the following specification under the assumption that the information nudge effects are homogeneous across the PG, AIPC, and control arms:

$$Y_{icgs} = \gamma H_c + \sum_{k \in \{\text{EN,BP,PR}\}} \left(\delta_k M^k_{icgs} + \theta_k H_c M^k_{icgs}\right) + \beta_1 T^{PG}_c + \beta_2 T^{AIPC}_c + \kappa X_{igs0} + \xi_{gs} + \varepsilon_{icgs}. \tag{3}$$

In this specification, $H_c$ is a class-level indicator for high-coverage information nudge assignment ($1 =$ high, $0 =$ low). The terms $M^k_{icgs}$ denote individual-level indicators for assignment to information nudge type $k \in \{\text{EN}, \text{BP}, \text{PR}\}$ (EN=children's expressed needs, BP=benefits of supportive parenting, PR=perspective-taking reflection), with the within-class control as the omitted group. $T^{PG}_c$ and $T^{AIPC}_c$ denote assignment to the parenting guidebook and AI-assisted parenting chatbot arms, as defined in Equations (1)–(2). $\xi_{gs}$ denotes strata (school-by-grade) fixed effects, and $X_{icgs0}$ is a vector of baseline covariates included to improve precision.

The coefficient $\delta_k$ captures the average effect of information nudge type $k$ in low-coverage classes ($H_c = 0$), while $\delta_k + \theta_k$ gives its effect in high-coverage classes ($H_c = 1$); thus, $\theta_k$ measures how information nudge effectiveness varies with coverage intensity.

When the impact of information nudges differs across various types of parenting guidance treatments, the fully saturated model is specified as follows:

$$
\begin{aligned}
Y_{icgs} = \alpha + \gamma H_c + & \sum_{k\in\{\text{EN,BP,PR}\}} \left(\delta_k M^k_{icgs} + \theta_k H_c M^k_{icgs}\right) \\
& + T^{PG}_c \Big[\beta_{\text{PG}} + \gamma_{\text{PG}} H_c + \sum_{k\in\{\text{EN,BP,PR}\}} \left(\delta^{\text{PG}}_k M^k_{icgs} + \theta^{\text{PG}}_k H_c M^k_{icgs}\right) \Big] \\
& + T^{AIPC}_c \Big[\beta_{\text{AIPC}} + \gamma_{\text{AIPC}} H_c + \sum_{k\in\{\text{EN,BP,PR}\}} \left(\delta^{\text{AIPC}}_k M^k_{icgs} + \theta^{\text{AIPC}}_k H_c M^k_{icgs}\right) \Big] \\
& + \kappa X_{icgs0} + \xi_{gs} + \epsilon_{icgs},
\end{aligned}
\tag{4}
$$

In this fully saturated specification, the baseline coefficients ($\gamma$, $\delta_k$, $\theta_k$) capture the effects of coverage intensity and information nudge assignment in the *between-class control* arm (in the main experimental arm). The terms multiplied by $T^{PG}_c$ and $T^{AIPC}_c$ represent *differences* in these effects relative to *between-class control* for the parenting guidebook and AI-assisted parenting chatbot arms, respectively. Specifically:

- $\delta_k$ is the effect of information nudge type $k$ in low-coverage *between-class control* classes; $\delta_k + \theta_k$ is its effect in high-coverage *between-class control* classes.

- $\delta_k + \delta^{\text{PG}}_k$ is the effect of information nudge $k$ in low-coverage PG treatment classes; $\delta_k + \delta^{\text{PG}}_k + \theta_k + \theta^{\text{PG}}_k$ is the effect in high-coverage PG classes (and analogously for AIPC).

- $\beta_{\text{PG}}$ and $\beta_{\text{AIPC}}$ capture the average effect of being assigned to the parenting guidebook or AI-assisted parenting chatbot arm (relative to *between-class control*) among parents who receive no information nudge and are in low-coverage classes.

- $\gamma$, $\gamma_{PG}$, and $\gamma_{AIPC}$ allow the pure effect of high-coverage assignment (e.g., via ambient exposure or spillovers) to differ across arms, even among non-recipients of information nudges.

### 5.1.4 Experimental demander effect

Administrative AI chatbot logs provide an objective safeguard against experimental demand effects, which often manifest as "social desirability bias" in self-reports. Unlike surveys, where parents may report "desired" behaviors to satisfy researcher expectations, logs capture actual engagement through metrics like message frequency, duration, and response timing. These real-time records are less prone to conscious bias, signaling genuine intervention use rather than superficial compliance.

Furthermore, a demand-effect interpretation fails to explain a strong "dosage-response" relationship between usage intensity and outcomes. If results were driven by reporting bias, improvements would likely appear uniformly across the treatment group regardless of actual engagement. Evidence that gains are concentrated among high-intensity users confirms a mechanism-driven impact, proving that the AI acts as a functional resource rather than an artifact of the experimental setting.

In a similar vein, we exploit the knowledge test on guidebook content as a diagnostic for PG treatment effect. If demand effects drive the PG treatment effect, parents who scored high on the test (i.e., those who actually read and comprehended the guidebook) and those who scored low (i.e., those who did not engage with the material) would exhibit similar treatment effects, since both groups are equally aware of their treatment status and would face the same incentive to report socially desirable responses. A finding that treatment effects are significantly larger among high-scoring parents indicates that outcomes improve through actual content absorption, instead of the demand effect.

### 5.1.5 Unintended Spillovers

The risk of unintended treatment spillovers to the between-class control group is minimized by both the nature of the intervention and the school structure. First, both the parenting guidebook and AI-chatbot guidance are "take-home" treatments. AI is delivered through a private digital platform, and its access is strictly controlled via administrative registration, ensuring that control group parents (who lack login credentials) cannot access the AI tool. While one might be concerned that behavioral improvements in treated parents could "spill over" to control parents through social interaction, this inter-class spillover is unlikely in our context. Students in different classrooms rarely interact due to the lack of shared inter-class courses or activities, and their parents are similarly socially distant. To empirically validate this assumption, we use classroom proximity as a robustness check. If spillovers were present, we would expect control classrooms located adjacent to treated classrooms to show different outcomes than those further away, a pattern we can test in larger school settings.

### 5.1.6 Peer spillovers to left-behind children

Beyond ruling out unintended spillover, our design enables us to formally estimate intentional spillovers to left-behind children. Specifically, because the intervention is directed at parents who present at home, we can examine how the treatment to parents of non-left-behind children affects their peers (left-behind children) within the same treatment environment. By comparing the outcomes of left-behind children in classrooms with varying densities of treated non-left-behind peers, we can quantify the indirect benefits of the intervention. This approach moves beyond simply dismissing spillover as a threat to internal validity; instead, it treats peer effects as a substantive mechanism of interest, allowing us to estimate the total social return of the AI intervention on the broader classroom ecosystem.

### 5.1.7 Missing data

To address the potential for bias introduced by sample loss, we document attrition and test for differential dropout rates across treatment arms. Should we detect evidence of non-random attrition, we apply inverse probability weights (IPW) to re-weight the observed sample based on baseline covariates, thereby restoring its representativeness relative to the initial population. Furthermore, we compute Lee bounds (Lee, 2009; Tauchmann, 2014), which establish the robust bounds for treatment effects by accounting for potential selection bias among the attriters. This dual strategy ensures that our findings remain internally valid and are not driven by differences in non-attrition.

### 5.1.8 Sample representativeness

Our data are collected from a rural county in Hunan Province. Thus, the representativeness of our survey data may be a concern, raising the issue of the external validity of our findings. To alleviate such concern, we compare our survey data with a nationally representative dataset, namely the China Family Panel Survey (CFPS). We will use rural children of the same age in the 2020 CFPS to compare the differences between our survey data and the CFPS data. We compare children's and parents' demographics to test whether these characteristics are small between the CFPS and our survey data.

We will follow Balboni et al. (2022) to conduct weighted regressions based on weights calculated from CFPS rural households to address the different sampling probabilities of households. Specifically, we categorize observations into different cells by crossing income deciles and mothers' ages. The frequencies of observations in each cell in the CFPS data are used as the estimation weights for the same "age-income" cell in our data. We will test whether the coefficients are similar to the main results.

### 5.1.9 Outliers

Outliers are not expected for survey-based outcomes, since these rely on validated questionnaires with bounded response scales. No winsorization is planned. Indices will be constructed following prespecified rules.

Test scores from administrative records may contain extreme observations, especially very low marks. We address this by standardizing scores (relative to the control group mean and SD), so estimated effects can be interpreted as standardized mean differences (effect size). As a robustness check, we also construct percentile ranks, defined as each student's empirical percentile position among peers. This reduces the influence of outliers and provides a clear measure of relative academic standing.

## 5.2 Heterogeneous Effects

We examine treatment effect heterogeneity along pre-specified baseline dimensions that reflect differences in children's needs and family environments. We begin with baseline child ability, including test scores, socioemotional skills, and mental health. The intervention emphasizes supportive parenting and emotional support, which are likely to matter most when children face challenges in these skills. We also examine heterogeneity by child gender, as parenting expectations and discipline practices differ for boys and girls in rural China, potentially shaping how changes in parenting translate into socio-emotional and academic outcomes.

We next consider heterogeneity by baseline parenting quality and by parental education. Parents with weaker baseline practices may have more scope to benefit from guidance, while parents whose practices are already well aligned may experience smaller gains. Education may further affect how parents interpret and implement advice, helping distinguish whether the intervention primarily improves parenting efficiency or compensates for gaps in parenting knowledge and skills.

Finally, we examine heterogeneity by baseline parent–child conflict intensity. Frequent conflict may indicate stress or ineffective communication, increasing the potential gains from guidance that emphasizes de-escalation and perspective-taking.

We examine heterogeneity using interaction models that interact each pre-determined characteristic with treatment indicators, controlling for baseline outcomes and covariates. For clarity, for some analyses, we also conduct subgroup analyses to complement the interaction models.

We also conduct exploratory analyses using machine-learning methods, focusing on generalized random forests (GRF). We use GRF to estimate conditional average treatment effects (CATEs) across a rich set of baseline characteristics, allowing for flexible, data-driven heterogeneity. We then project the estimated CATEs onto the pre-specified moderators to assess consistency with the interaction models and to visualize the distribution of predicted effects. Together, the interaction regressions, subgroup ITT plots, and exploratory GRF analyses provide a comprehensive view of both policy-relevant and unanticipated heterogeneity.

## 5.3 Multiple hypothesis testing

We adopt a hierarchical approach to multiple-hypothesis testing (MHT). Specifically, we organize outcomes into three tiers (see Table 1):

- *Intermediate outcomes* — measures organized into two outcome families: (a) parenting style, practices, and parent–child relationship quality; (b) AI chatbot and guidebook usage and parental educational inputs.
- *Main outcomes* — student mental health and academic achievement, treated as a single outcome family.
- *Other outcomes* — organized into two outcome families: (a) socio-emotional skills, non-cognitive skills, and creativity; (b) peer networks.

Within each outcome family, we construct standardized indices that aggregate related measures into a single summary statistic, following Anderson (2008). For each tier, we also form an overall index to capture broader program effects. We also adjust $p$-values using the stepdown procedure of Romano and Wolf (2005).

# 6 Interpretation of the results

## 6.1 Interpreting Alternative Findings

While several outcome patterns are possible, our prior is that the AI-assisted guidance outperforms both the booklet and the control. Such a pattern would be consistent with theories emphasizing implementation frictions rather than information deficits. Both treatments deliver similar parenting knowledge, but the AI intervention adds personalization, interactivity, and timely prompts, which may help parents translate knowledge into practice, particularly when parenting challenges vary across children and situations. Stronger effects of the AI arm would therefore point to personalization and just-in-time support as key mechanisms for improving parenting behavior.

If we instead observe null effects or negative impacts, the rich set of intermediate outcomes allows us to diagnose where the intervention breaks down. For example, improvements in parenting knowledge without corresponding changes in practices would indicate implementation failures rather than informational gaps. Changes in reported practices without improvements in parent–child relationships or child outcomes would suggest limited transmission from parenting behavior to child development. Alternatively, increases in parent–child conflict or parental stress would point to unintended consequences, such as misapplied guidance or time crowd-out. By tracing effects along the program's theory of change, we can distinguish between failures of take-up, implementation, and downstream impact, and thereby inform the redesign of future parenting interventions.

## 6.2 Policy Implications

Policy implications depend on the relative performance of the interventions. If the AI-assisted guidance outperforms the booklet and both improve outcomes, the next step is to assess cost-effectiveness and consider scaling AI delivery through schools, conditional on affordability and sustained take-up. If the booklet performs similarly or better, it provides a low-cost option suitable for broad deployment.

If the interventions yield null effects, policymakers should reassess take-up and engagement, including the role of reminder nudges. If instrumental-variable estimates suggest meaningful effects but compliance is low, usage incentives may be warranted. If impacts emerge on parenting practices but not on child outcomes, longer follow-up may be needed to capture delayed effects. Any evidence of harm would argue against scale-up and motivate redesign with appropriate safeguards.

Heterogeneity results further inform implementation. Examining the distribution of CATEs suggests scope for targeting interventions toward families with the highest expected returns, particularly under budget constraints. We complement this with a cost-effectiveness analysis that accounts for usage-dependent AI costs, in contrast to the largely fixed costs of the booklet, allowing policymakers to align scale and targeting with available resources.

# References


**Alam, Mafruha, Ian B Hickie, Adam Poulsen, Mahalakshmi Ekambareshwar, Victoria Loblay, Jacob Crouse, Gabrielle Hindmarsh, Yun JC Song, Adam Yoon, Grace Cha et al.**, "Parenting app to support socio-emotional and cognitive development in early childhood: iterative codesign learnings from nine low-income and middle-income countries," *BMJ open*, 2023, *13* (5), e071232.

**Almond, Douglas, Janet Currie, and Valentina Duque**, "Childhood circumstances and adult outcomes: Act II," *Journal of Economic Literature*, 2018, *56* (4), 1360–1446.

**Anderson, Michael L**, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 2008, *103* (484), 1481–1495.

**Andresen, Elena M, Judith A Malmgren, William B Carter, and Donald L Patrick**, "Screening for depression in well older adults: The CES-D 10," *Journal of the American Geriatrics Society*, 1994, *42* (12), 1261–1268.

**Attanasio, Orazio, Darwin Cortes, Dario Maldonado, Paul Rodriguez-Lesmes, Nathalie Charpak, Rejean Tessier, Juan G Ruiz, Juan Gallego, Tiberio Hernandez, Felipe Uriza et al.**, "Parental Investments and Skills Formation During Infancy and Youth: Long Term Evidence From an Early Childhood Intervention," Technical Report, National Bureau of Economic Research 2024.

**Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil**, "Why Do People Stay Poor?," *The Quarterly Journal of Economics*, 2022, *137* (2), 785–844.

**Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Ozge Kabakcı, and Rei Mariman**, "Generative ai can harm learning," *Available at SSRN*, 2024, *4895486*.

**Baumrind, Diana**, "The influence of parenting style on adolescent competence and substance use," in "Adolescents and Their Families," Routledge, 2013, pp. 22–61.

**Carneiro, Pedro and James J Heckman**, "Human capital policy," *NBER Working Paper No. 9495*, 2003.

**_ , Emanuela Galasso, Italo Lopez Garcia, Paula Bedregal, and Miguel Cordero**, "Impacts of a large-scale parenting program: Experimental evidence from Chile," *Journal of Political Economy*, 2024, *132* (4), 1113–1161.

**Chua, Amy**, *Battle hymn of the tiger mother*, Bloomsbury publishing, 2011.

**Cunha, Flavio, Qinyou Hu, Yiming Xia, and Naibao Zhao**, "Reducing bullying: Evidence from a parental involvement program on empathy education," Technical Report, National Bureau of Economic Research 2023.

**Currie, Janet**, "Presidential Address: Investing in Children to Address the Child Mental Health Crisis," *American Economic Review*, 2025, *115* (5), 1369–1399.

**Doepke, Matthias and Fabrizio Zilibotti**, "Parenting with style: Altruism and paternalism in intergenerational preference transmission," *Econometrica*, 2017, *85* (5), 1331–1371.

**Dweck, Carol S**, *Mindset: The new psychology of success*, Random House, 2006.

**Edlund, Lena**, "A Review of Matthias Doepke and Fabrizio Zilibotti's Love, Money and Parenting: How Economics Explains the Way We Raise Our Kids," *Journal of Economic Literature*, 2021, *59* (3), 1001–1022.

**Fan, Weihua and Michael Chen**, "Parental involvement and students' academic achievement: A meta-analysis," *Educational Psychology Review*, 2001, *13* (1), 1–22.

**Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M Chang, and Sally Grantham-McGregor**, "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 2014, *344* (6187), 998–1001.

**Goodman, Robert**, "The Strengths and Difficulties Questionnaire: A research note," *Journal of Child Psychology and Psychiatry*, 1997, *38* (5), 581–586.

**Gosling, Samuel D, Peter J Rentfrow, and William B Swann**, "A Very Brief Measure of the Big-Five Personality Domains," *Journal of Research in Personality*, 2003, *37* (6), 504–528.

**Heckman, James J, Seong Hyeok Moon, Rodrigo Pinto, Peter A Savelyev, and Adam Yavitz**, "The rate of return to the HighScope Perry Preschool Program," *Journal of Public Economics*, 2010, *94* (1-2), 114–128.

**Justino, Patricia, Marinella Leone, Pierfrancesco Rolla, Monique Abimpaye, Caroline Dusabe, Marie D Uwamahoro, and Richard Germond**, "Improving parenting practices for early child development: Experimental evidence from Rwanda," *Journal of the European Economic Association*, 2023, *21* (4), 1510–1550.

**Lee, David S.**, "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 2009, *76* (3), 1071–1102.

**Maccoby, Eleanor E. and John A. Martin**, "Socialization in the context of the family: Parent-child interaction," in Paul H. Mussen and E. Mavis Hetherington, eds., *Handbook of Child Psychology: Vol. 4. Socialization, personality, and social development*, 4th ed., New York: John Wiley & Sons, 1983, pp. 1–101.

**Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian**, "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 2019, *109* (4), 1426–1460.

**Rammstedt, Beatrice and Oliver P John**, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, 2007, *41* (1), 203–212.

**Repetti, Rena L, Shelley E Taylor, and Teresa E Seeman**, "Risky families: family social environments and the mental and physical health of offspring.," *Psychological bulletin*, 2002, *128* (2), 330.

**Romano, Joseph P and Michael Wolf**, "Stepwise multiple testing as formalized data snooping," *Econometrica*, 2005, *73* (4), 1237–1282.

**Simone, Martin Elias De, Federico Hernan Tiberti, Maria Rebeca Barron Rodriguez, Federico Alfredo Manolio, Wuraola Mosuro, and Eliot Jolomi Dikoru**, "From chalkboards to chatbots: Evaluating the impact of generative AI on learning outcomes in nigeria," Technical Report, The World Bank 2025.

**Steinberg, Laurence**, "We know some things: Parent–adolescent relationships in retrospect and prospect," *Journal of research on adolescence*, 2001, *11* (1), 1–19.

**Tauchmann, Harald**, "Lee (2009) treatment-effect bounds for nonrandom sample selection," *The Stata Journal*, 2014, *14* (4), 884–894.

**UNICEF**, "Nearly 400 Million Young Children Worldwide Regularly Experience Violent Discipline at Home," https://www.unicef.org/press-releases/nearly-400-million-young-children-worldwide-regularly-experience-violent-discipline June 2024. Press release. Accessed: 12 February 2026.

**Ward, Catherine L, Inge M Wessels, Jamie M Lachman, Judy Hutchings, Lucie D Cluver, and Reshma Kassanjee**, "Parenting for lifelong health for young children: A randomized controlled trial in South Africa," *Journal of Child Psychology and Psychiatry*, 2020, *61* (4), 503–512.

**York, Benjamin N, Susanna Loeb, and Christopher Doss**, "One step at a time: The effects of an early literacy text-messaging program for parents of preschoolers," *Journal of Human Resources*, 2019, *54* (3), 537–566.

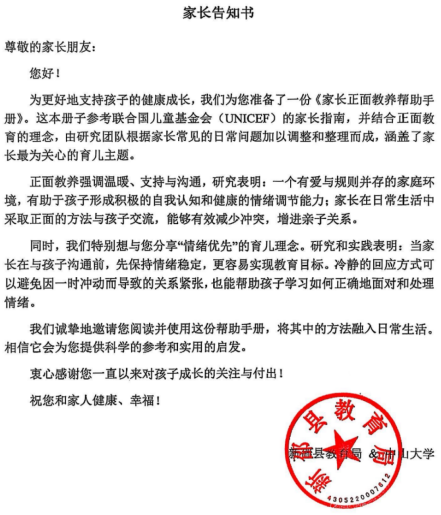
# 7 Administrative information

**Funding:** This work was supported by the National Science Foundation of China (NSFC) [Nos.72573191 and 72503195].

**Institutional Review Board approval**: This study was approved by the Research Ethics Committee of the China Center for Behavioral Economics and Finance (CCBEF), IRB No. 2025B05.

**Declaration of interest:** None.

# Figures and Tables

家长告知书

尊敬的家长朋友：

您好！

为更好地支持孩子的健康成长，我们为您准备了一份《家长正面教养帮助手册》。这本册子参考联合国儿童基金会（UNICEF）的家长指南，并结合正面教育的理念，由研究团队根据家长常见的日常问题加以调整和整理而成，涵盖了家长最为关心的育儿主题。

正面教养强调温暖、支持与沟通，研究表明：一个有爱与规则并存的家庭环境，有助于孩子形成积极的自我认知和健康的情绪调节能力；家长在日常生活中采取正面的方法与孩子交流，能够有效减少冲突，增进亲子关系。

同时，我们特别想与您分享"情绪优先"的育儿理念。研究和实践表明：当家长在与孩子沟通前，先保持情绪稳定，更容易实现教育目标。冷静的回应方式可以避免因一时冲动而导致的关系紧张，也能帮助孩子学习如何正确地面对和处理情绪。

我们诚挚地邀请您阅读并使用这份帮助手册，将其中的方法融入日常生活。相信它会为您提供科学的参考和实用的启发。

衷心感谢您一直以来对孩子成长的关注与付出！

祝您和家人健康、幸福！

新邵县教育局 & 中山大学

Figure 1: Cover of the Parenting Guidebook and Letter from Local Education Bureau

*Notes:* This figure presents the cover page of our designed parenting guidebook and the letter from the local Education Bureau. The guidebook, adapted from UNICEF's parenting education guidelines and tailored to the Chinese cultural context, includes practical examples related to ten key issues identified by local parents in a preliminary survey.

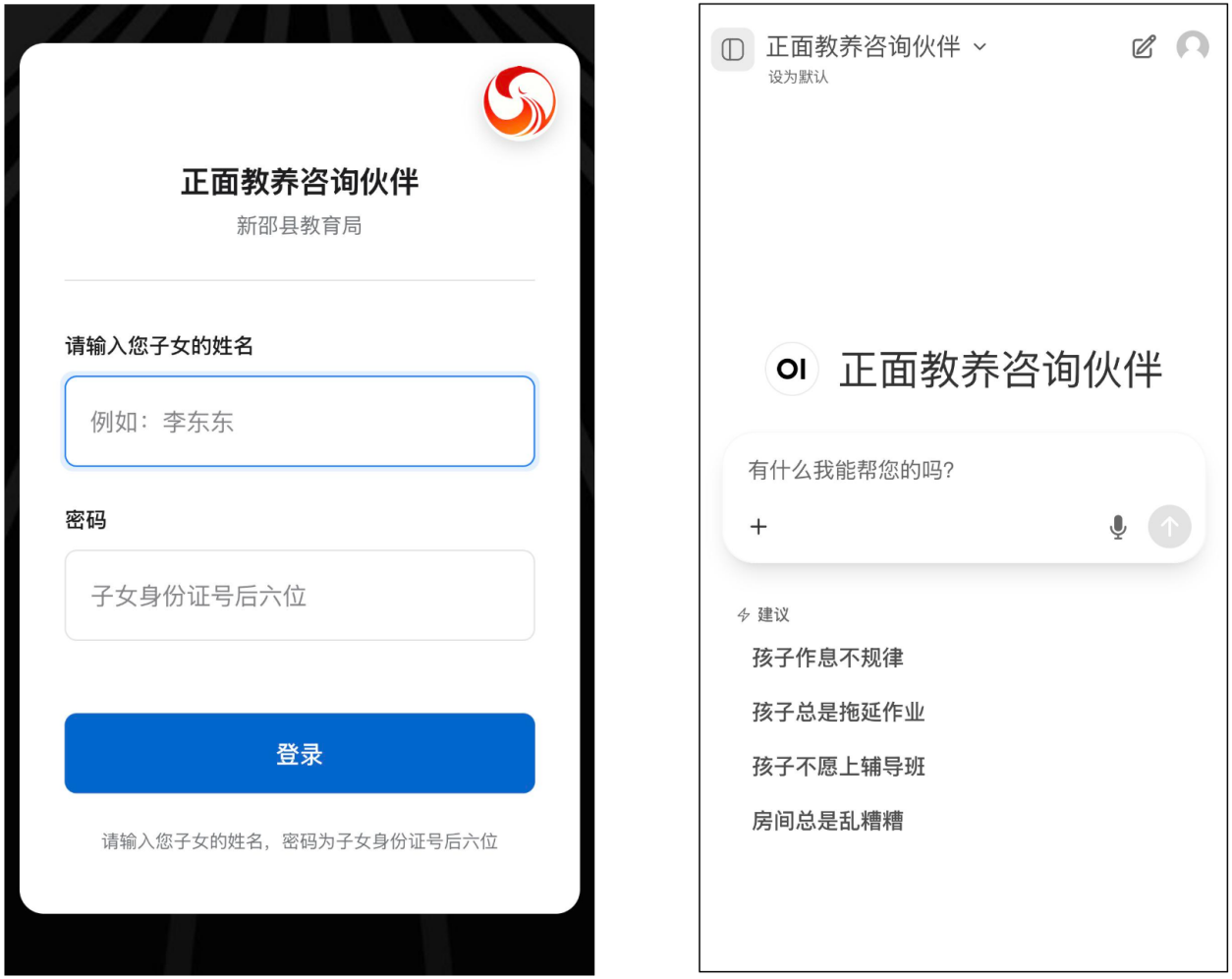



Figure 2: Login Page of the AI-Assisted Parenting Chatbot App

*Notes:* This figure presents the login page of our designed app of an AI-assisted parenting chatbot, which is based on a trained large language model (Deepseek).

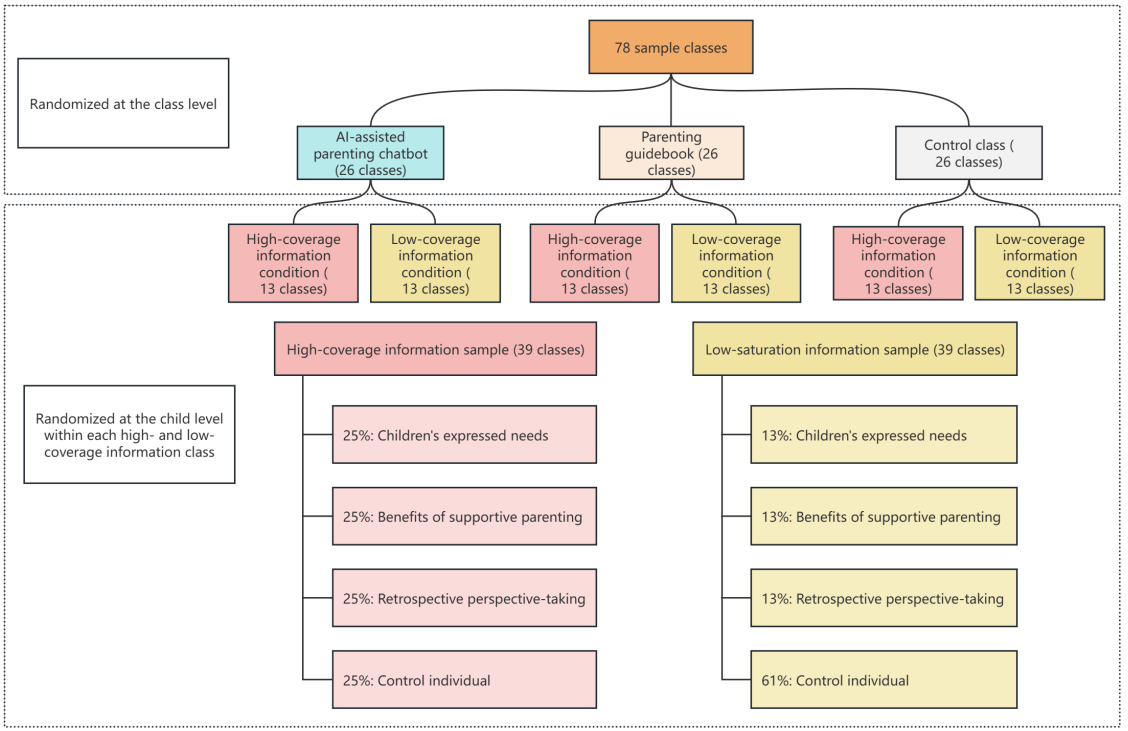



Figure 3: Experiment Design

*Notes:* The figure presents the experimental design, which incorporates two levels of randomization. (i) Across classes: Seventy-eight sample classes were randomized into three arms: 26 classes were assigned to the AIPC treatment arm, 26 to the PG treatment arm, and 26 to the no-guidance arm (control). Within each arm, classes were further divided into 13 high-coverage information groups and 13 low-coverage information groups. (ii) Across individuals: In the high-coverage information sample, children were equally distributed across four conditions: 25% of parents received the children's expressed needs message, 25% received the benefits of supportive parenting message, 25% received the retrospective perspective-taking reflection message, and 25% served as individual controls. In the low-coverage information sample, the distribution was unequal: 13% each were assigned to receive the children's expressed needs message, the benefits of supportive parenting message, and the perspective-taking reflection message, while 61% of parents served as controls. Cells shaded in pink represent the high-coverage sample, and cells shaded in yellow represent the low-coverage sample.

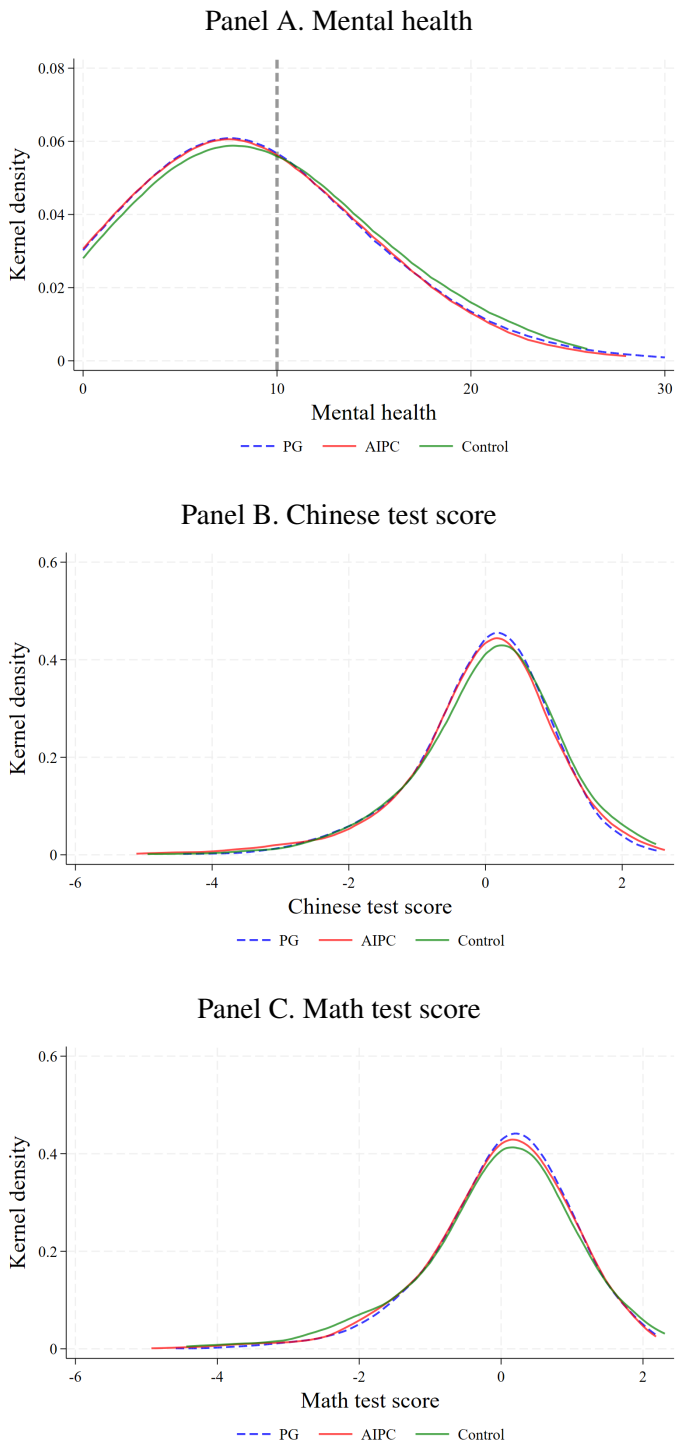



Figure 4: Distribution of Children's Mental Health, Chinese and Math Test Scores

*Notes:* The figure displays the distribution of mental health, Chinese, and math test scores for children in the PG, AIPC, and control classes measured in the baseline survey. Children are those with at least one parent at home (non-left-behind children). Mental health is assessed using the CES-D 10 scale, with higher scores indicating a higher level of stress. Test scores are standardized within each school-by-grade cell to have a mean of 0 and a standard deviation of 1.

Table 1: Outcomes and Measurements in Child and Parent Survey

| Outcome | Child Survey | Parent Survey |
|---|---|---|
| **Intermediate Outcomes** | | |
| 1. Parenting style and practices | Parenting style (Maccoby and Martin, 1983) (C.1.14) | Discipline strategies, Rule enforcement, Parent empathy, Positive parenting, Perceived efficacy of different parenting practices (C.2.5-C.2.9) |
| 2. Parent–child relationship quality | # of verbal and physical punishment (C.1.3) | # of conflicts with children (C.2.4) |
| 3. AI chatbot and guidebook usage | – | Chatbot usage record (app record), Test for knowledge in booklet, perception on AI in parenting (C.2.2) |
| 4. Parental educational inputs | Parental time input (C.1.15) | Parental educational expenditure |
| **Main Outcomes** | | |
| 1. Mental health | Children's CES-D 10 (Andresen et al., 1994) (C.1.4) | Parental CES-D 10 (Andresen et al., 1994) (C.2.3) |
| 2. Academic outcomes | Chinese and mathematics test scores | Satisfaction with performance, Grade expectations (C.2.1) |
| **Other Outcomes** | | |
| 1. Socio-emotional skills | SDQ domains (Goodman, 1997) (C.1.6) | – |
| 2. Non-cognitive skills | Economic Preference (Delay of gratification, Risk preference, Altruism), Big Five (BFI-10), Self-control, Misbehavior, Growth mindset (Rammstedt and John, 2007; Dweck, 2006) (C.1.1,C.1.2,C.1.5,C.1.7,C.1.8,C.1.9, C.1.11) | – |
| 3. Creativity | Self-perceived creativity (C.1.10) | – |
| 4. Peer networks | Number of friends and perceived peer relationship quality(C.1.12,C.1.13) | – |

*Notes*: This table shows the intermediate, main, and other outcomes, which are organized as families of outcomes. The related item questions are labeled in parentheses, which can be found in the Appendix Section C.

Table 2: Summary Statistics of Variables from Children's Reports (Non-Left-Behind Children Sample)

| | Obs. (1) | Mean (2) | S.D. (3) | Min (4) | Max (5) |
|---|---|---|---|---|---|
| *Panel A. Children's characteristic* | | | | | |
| Boy (1=yes) | 2,505 | 0.531 | 0.499 | 0 | 1 |
| Age | 2,505 | 9.848 | 0.643 | 9 | 12 |
| Rural *hukou* (1=yes) | 2,505 | 0.724 | 0.447 | 0 | 1 |
| Non-minority (1=yes) | 2,505 | 0.978 | 0.147 | 0 | 1 |
| # of siblings | 2,505 | 1.457 | 1.039 | 0 | 5 |
| # of illnesses (last semester) | 2,505 | 1.513 | 1.468 | 0 | 7 |
| *Panel B. Children's outcomes* | | | | | |
| Chinese test score | 2,505 | 77.675 | 9.18 | 35 | 97 |
| Math test score | 2,505 | 77.855 | 12.029 | 27 | 100 |
| Mental health (CES-D 10) | 2,505 | 8.672 | 5.273 | 0 | 30 |
| Delay of gratification | 2,505 | 0.63 | 0.483 | 0 | 1 |
| Risk preference | 2,505 | 0.282 | 0.45 | 0 | 1 |
| Altruism | 2,505 | 3.642 | 1.481 | 0 | 8 |
| SDQ: | | | | | |
| Prosocial behavior | 2,505 | 4.685 | 1.094 | 2 | 6 |
| Internalizing behavior problems | 2,505 | 9.059 | 1.697 | 4 | 12 |
| Externalizing behavior problems | 2,505 | 9.461 | 1.663 | 4 | 12 |
| Big-Five: | | | | | |
| Extroversion | 2,505 | 8.889 | 2.548 | 2 | 14 |
| Agreeableness | 2,505 | 9.625 | 2.394 | 2 | 14 |
| Conscientiousness | 2,505 | 7.821 | 2.618 | 2 | 14 |
| Neuroticism | 2,505 | 8.630 | 2.714 | 2 | 14 |
| Openness | 2,505 | 8.533 | 2.518 | 2 | 14 |
| Growth mindest | 2,505 | 11.58 | 2.897 | 3 | 15 |
| Creativity | 2,505 | 26.455 | 3.878 | 14 | 36 |
| Self-control | 2,505 | 11.734 | 3.235 | 4 | 20 |
| # of misbehaviors (last semester) | 2,505 | 30.911 | 4.307 | 14 | 40 |
| # of nominated friends | 2,505 | 5.183 | 2.496 | 0 | 8 |
| Perceived current peer relationship quality | 2,505 | 3.515 | 0.978 | 0 | 4 |
| # of parents' verbal abuse and physical punishment (last semester) | 2,505 | 7.836 | 7.217 | 0 | 28 |
| Need to express oneself to parents | 2,505 | 2.926 | 1.055 | 1 | 4 |
| Perceived parental misunderstanding | 2,505 | 2.161 | 1.126 | 1 | 4 |
| Need for parental listening | 2,505 | 2.919 | 1.019 | 1 | 4 |

*Continued on next page*

*(Continued) Summary Statistics*

| | Obs. | Mean | S.D. | Min | Max |
|---|---|---|---|---|---|
| *Panel C. Parents' parenting style* | | | | | |
| Father: | | | | | |
| Authoritative | 2,505 | 10.535 | 3.043 | 3 | 15 |
| Authoritarian | 2,505 | 6.077 | 3.125 | 3 | 15 |
| Permissive | 2,505 | 7.951 | 2.988 | 3 | 15 |
| Neglectful | 2,505 | 4.657 | 2.228 | 2 | 10 |
| Mother: | | | | | |
| Authoritative | 2,505 | 11.074 | 3.049 | 3 | 15 |
| Authoritarian | 2,505 | 6.178 | 3.025 | 3 | 15 |
| Permissive | 2,505 | 8.004 | 3.041 | 3 | 15 |
| Neglectful | 2,505 | 4.638 | 1.747 | 2 | 10 |
| *Panel D: Child's daily time spent (in minutes)* | | | | | |
| Doing homework, reviewing, and previewing lessons | 2,505 | 45.637 | 30.786 | 0 | 120 |
| Extracurricular reading | 2,505 | 25.191 | 21.662 | 0 | 120 |
| Using Internet and playing video games | 2,505 | 20.743 | 24.169 | 0 | 120 |
| Watching TV | 2,505 | 26.628 | 26.647 | 0 | 120 |
| Doing sports/physical activities | 2,505 | 29.637 | 27.275 | 0 | 120 |
| Playing with friends/classmates | 2,505 | 37.92 | 30.461 | 0 | 120 |
| Playing alone | 2,505 | 28.317 | 28.929 | 0 | 120 |
| Studying with parental supervision | 2,505 | 31.457 | 30.188 | 0 | 120 |
| Playing with parents | 2,505 | 17.727 | 25.448 | 0 | 120 |
| Listening to stories told by parents | 2,505 | 20.313 | 24.495 | 0 | 120 |

*Notes*: *Delay of gratification* is a binary variable equal to 1 if the child chooses the delayed payment and 0 if the child chooses the immediate payment. *Risk preference* is a binary variable equal to 1 if the child chooses to flip a coin (RMB 10 if heads and RMB 0 if tails), and 0 if the child chooses to receive RMB 5 for sure. Mental health is measured using the CES-D 10 scale and is constructed by summing scores across the ten items; a higher total score indicates more severe depressive symptoms. *Altruism* is measured using a dictator game and equals the number of candies (0–8) the child chooses to give to an anonymous classmate; higher values indicate greater altruism. *Self-control* is an index constructed from four Likert-scale items on planning, time management, impulsivity, and acting without thinking. Higher values indicate higher self-control. *Prosocial behavior* is an index capturing empathy and willingness to help others; higher values indicate more prosocial behavior. *Internalizing behavior problems* is constructed as the sum of the emotional and peer-problem components. *Externalizing behavior problems* is constructed as the sum of the conduct and attention components. Higher values indicate more severe behavior problems. *Growth mindset* is a continuous index constructed by summing responses to seven 5-point Likert-scale items; higher values indicate a stronger belief that abilities (e.g., intelligence) can be improved through effort. *Creativity* is a continuous index constructed by summing responses to nine 4-point Likert-scale items; higher values indicate greater creativity. See Section C.1 for details.

Table 3: Summary Statistics of Variables from Parents' Reports (Non-Left-Behind Children Sample)

| | Obs. (1) | Mean (2) | S.D. (3) | Min (4) | Max (5) |
|---|---|---|---|---|---|
| *Panel A. Parents' characteristics* | | | | | |
| Father's age | 2,505 | 39.685 | 4.576 | 30 | 56 |
| Mother's age | 2,505 | 37.225 | 4.48 | 28 | 52 |
| Father' years of schooling | 2,505 | 11.539 | 3.329 | 3 | 19 |
| Mother' years of schooling | 2,505 | 11.456 | 3.371 | 3 | 19 |
| Father employed in the local county (1=yes) | 2,505 | 0.569 | 0.495 | 0 | 1 |
| Mother employed in the local county (1=yes) | 2,505 | 0.916 | 0.277 | 0 | 1 |
| *Panel B. Parents' AI usage and perceived role in parenting* | | | | | |
| Using AI for child education (1=yes) | 2,505 | 0.063 | 0.244 | 0 | 1 |
| Using AI for own study/work (1=yes) | 2,505 | 0.271 | 0.445 | 0 | 1 |
| Using AI for other purposes (1=yes) | 2,505 | 0.118 | 0.322 | 0 | 1 |
| Perceived role of AI in parenting | 2,505 | 19.089 | 5.223 | 4 | 28 |
| *Panel C. Parents' evaluation of their children* | | | | | |
| Satisfaction with the child's academic performance in the last semester | 2,505 | 7.112 | 1.868 | 0 | 10 |
| Rating of the child's academic performance in the last semester | 2,505 | 76.421 | 19.056 | 8 | 100 |
| Expectation for the child's academic achievement | 2,505 | 2.675 | 0.502 | 1 | 3 |
| *Panel D. Parents' outcomes and their parenting practices* | | | | | |
| Parents' mental health (CES-D 10) | 2,505 | 8.145 | 4.247 | 0 | 27 |
| # of parent–child conflicts (last semester) | 2,505 | 18.028 | 14.249 | 0 | 70 |
| Parenting skills in the last semester: | | | | | |
| Rule enforcement | 2,505 | 8.145 | 1.408 | 2 | 10 |
| Emotional regulation | 2,505 | 7.442 | 1.383 | 2 | 10 |
| Communication skills | 2,505 | 7.874 | 1.648 | 2 | 10 |
| Positive parenting practices | 2,505 | 15.53 | 2.954 | 4 | 20 |
| Parenting self-efficacy | 2,505 | 21.746 | 4.104 | 6 | 30 |
| Parenting awareness | 2,505 | 30.946 | 7.471 | 8 | 40 |
| Punitive discipline measures | 2,505 | 0.998 | 1.534 | 0 | 9 |
| Command-style discipline measures | 2,505 | 1.256 | 1.674 | 0 | 9 |
| Permissive discipline measures | 2,505 | 0.319 | 0.877 | 0 | 9 |
| Communicative discipline measures | 2,505 | 6.45 | 2.625 | 0 | 9 |
| Empathy | 2,505 | 14.395 | 2.867 | 4 | 20 |
| *Panel E. Monthly Expenditures and income (in CNY)* | | | | | |
| Monthly spending on children's school supplies | 2,505 | 69.771 | 74.208 | 0 | 300 |
| Monthly spending on extracurricular resources | 2,505 | 75.544 | 75.761 | 0 | 300 |
| Monthly spending on toys | 2,505 | 53.853 | 73.898 | 0 | 300 |
| Monthly household income | 2,505 | 6,865.40 | 3,794.03 | 250 | 14,000 |

*Continued on next page*

*(Continued) Summary Statistics*

| | Obs. | Mean | S.D. | Min | Max |
|---|---|---|---|---|---|
| *Panel F. Paternal and maternal grandparents' parenting styles during the parents' childhoods* | | | | | |
| Father's grandparents: | | | | | |
| Authoritative | 2,505 | 6.391 | 1.577 | 3 | 9 |
| Authoritarian | 2,505 | 4.475 | 1.249 | 2 | 6 |
| Mother's grandparents: | | | | | |
| Authoritative | 2,505 | 6.734 | 1.566 | 3 | 9 |
| Authoritarian | 2,505 | 4.186 | 1.269 | 2 | 6 |

*Notes*: *Mental health* is measured using the CES-D 10 scale and is constructed by summing scores across the ten items; a higher total score indicates more severe depressive symptoms. *Perceived role of AI in parenting* is an index constructed by summing four 7-point Likert-scale items on parents' attitudes toward AI tools in parenting; higher values indicate a more positive perceived role of AI in communication, understanding the child, educational advice, and parent–child relationship quality. *Parenting skills* are measured using six 5-point scenario-based items and summarized into three indices—*rule enforcement*, *emotional regulation*, and *communication skills*—each constructed by summing two items. All indices are coded so that higher values indicate better parenting skills in the corresponding dimension. *Positive parenting practices* is an index constructed by summing four 5-point Likert-scale items on supportive and emotionally responsive parenting behaviors; higher values indicate more positive parenting practices. *Parenting self-efficacy* is measured as the sum of six 5-point Likert-scale items; higher values indicate greater confidence and satisfaction in the parenting role. *Parenting awareness* is measured as the sum of eight 5-point Likert-scale items; higher values indicate greater awareness of effective parenting and the potential harms of harsh discipline and overindulgence. *Punitive discipline measure*, *command-style discipline measure*, *permissive discipline measure*, and *communicative discipline measure* are constructed from nine discipline vignettes and equal the number of situations (0–9) in which the parent reports using the corresponding practice; higher values indicate more frequent use of that practice. See Section C.2 for details.

Table 4: Correlations between parenting styles and children's academic performance and psychological stress.

| | Chinese test score | Math test score | CES-D score |
|---|---|---|---|
| | (1) | (2) | (3) |
| Authoritative (father) | 0.012 | 0.053** | -0.125*** |
| | (0.024) | (0.024) | (0.022) |
| Authoritarian (father) | 0.083*** | 0.056** | 0.109*** |
| | (0.023) | (0.024) | (0.022) |
| Permissive (father) | 0.033 | 0.025 | -0.042* |
| | (0.024) | (0.024) | (0.022) |
| Neglectful (father) | -0.050** | -0.014 | 0.053** |
| | (0.023) | (0.023) | (0.021) |
| Authoritative (mother) | 0.058** | 0.032 | -0.134*** |
| | (0.024) | (0.024) | (0.022) |
| Authoritarian (mother) | 0.057** | 0.061*** | 0.116*** |
| | (0.023) | (0.023) | (0.021) |
| Permissive (mother) | -0.126*** | -0.131*** | -0.062** |
| | (0.028) | (0.028) | (0.026) |
| Neglectful (mother) | -0.095*** | -0.099*** | 0.080*** |
| | (0.025) | (0.025) | (0.024) |
| Observations | 2,505 | 2,505 | 2,505 |
| Class FE | Yes | Yes | Yes |

*Notes*: This table reports the correlations between parenting styles and children's academic performance and psychological stress. We regress students' test scores and stress outcomes on fathers' and mothers' parenting styles, including class fixed effects. Test scores are standardized within each school-by-grade cell to have a mean of 0 and a standard deviation of 1, while psychological stress and parenting-style measures are standardized using the full sample. Standard errors clustered at the class level are reported in parentheses. *, **, and *** represent the significance levels of 10%, 5%, and 1%, respectively.

Table 5: Balance test for Variables From Children's Report (Non-Left-Behind Children)

| Variable | AIPC vs. Control | PG vs. Control | High-coverage vs. Low-coverage | Children's expressed needs vs. No message | Benefits of supportive parenting vs. No message | Perspective-taking reflection vs. No message |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A. Children's characteristics* | | | | | | |
| Boy | 0.008 | 0.010 | -0.009 | 0.021 | 0.048* | 0.042 |
| | (0.017) | (0.015) | (0.015) | (0.028) | (0.028) | (0.028) |
| Age | -0.052 | -0.078** | -0.122 | -0.043 | -0.002 | -0.010 |
| | (0.036) | (0.031) | (0.099) | (0.033) | (0.027) | (0.031) |
| Rural | 0.017 | 0.007 | -0.013 | -0.001 | -0.031 | -0.017 |
| | (0.035) | (0.037) | (0.043) | (0.025) | (0.024) | (0.026) |
| Non-minority | 0.003 | 0.006 | -0.002 | 0.007 | 0.004 | 0.002 |
| | (0.007) | (0.006) | (0.007) | (0.008) | (0.011) | (0.009) |
| # of siblings | -0.053 | -0.036 | -0.031 | 0.039 | 0.115 | -0.022 |
| | (0.075) | (0.074) | (0.074) | (0.056) | (0.074) | (0.052) |
| # of illnesses during the last semester | 0.092 | 0.029 | -0.034 | 0.011 | 0.051 | -0.081 |
| | (0.064) | (0.063) | (0.068) | (0.077) | (0.089) | (0.078) |
| *Panel B. Children's outcomes* | | | | | | |
| Chinese test score | 0.062 | -0.025 | 0.013 | 0.029 | -0.098 | -0.068 |
| | (0.064) | (0.066) | (0.054) | (0.055) | (0.065) | (0.062) |
| Math test score | -0.069 | -0.042 | -0.066 | 0.024 | 0.073 | 0.065 |
| | (0.067) | (0.053) | (0.051) | (0.057) | (0.058) | (0.055) |
| Mental health | 0.388 | -0.207 | 0.087 | 0.250 | 0.169 | 0.095 |
| | (0.334) | (0.299) | (0.356) | (0.316) | (0.297) | (0.311) |
| Delay of gratification | 0.012 | -0.017 | -0.028 | -0.009 | -0.018 | 0.014 |
| | (0.032) | (0.034) | (0.030) | (0.026) | (0.024) | (0.028) |
| Altruism | 0.034 | 0.017 | -0.006 | 0.074 | 0.043 | 0.028 |
| | (0.103) | (0.084) | (0.097) | (0.089) | (0.082) | (0.091) |
| Risk preference | -0.021 | 0.002 | 0.000 | 0.035 | -0.020 | 0.020 |
| | (0.026) | (0.025) | (0.022) | (0.026) | (0.023) | (0.034) |
| Sefl-control | 0.110 | 0.150 | 0.236 | -0.045 | -0.169 | -0.162 |
| | (0.206) | (0.201) | (0.205) | (0.216) | (0.163) | (0.201) |
| SDQ: | | | | | | |
| Prosocial behavior | 0.008 | -0.043 | 0.028 | -0.017 | -0.069 | -0.051 |
| | (0.072) | (0.074) | (0.072) | (0.058) | (0.063) | (0.059) |
| Internalizing behavior problems | -0.008 | 0.015 | 0.040 | 0.070 | -0.076 | -0.011 |

*Continued on next page*

*(Continued) Balance test*

| Variable | AIPC vs. Control | PG vs. Control | High-coverage vs. Low-coverage | Children's needs message vs. No message | Positive effects message vs. No message | Perspective-taking message vs. No message |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | (0.105) | (0.084) | (0.086) | (0.111) | (0.086) | (0.099) |
| Externalizing behavior problems | -0.009 | 0.048 | 0.070 | -0.078 | -0.122 | -0.080 |
| | (0.114) | (0.111) | (0.100) | (0.097) | (0.091) | (0.100) |
| Big-five: | | | | | | |
| Extroversion | -0.059 | 0.000 | -0.158 | 0.116 | -0.038 | -0.167 |
| | (0.142) | (0.130) | (0.141) | (0.142) | (0.155) | (0.150) |
| Agreeableness | 0.246** | 0.107 | 0.060 | -0.067 | -0.076 | -0.059 |
| | (0.116) | (0.140) | (0.127) | (0.152) | (0.129) | (0.126) |
| Conscientiousness | 0.103 | 0.098 | -0.088 | -0.085 | -0.076 | -0.102 |
| | (0.154) | (0.171) | (0.145) | (0.148) | (0.150) | (0.162) |
| Neuroticism | -0.103 | -0.123 | 0.006 | -0.124 | -0.001 | -0.197 |
| | (0.140) | (0.137) | (0.130) | (0.156) | (0.176) | (0.175) |
| Openness | -0.025 | -0.102 | -0.197 | -0.269* | -0.189 | -0.102 |
| | (0.127) | (0.122) | (0.120) | (0.157) | (0.135) | (0.145) |
| Growth mindest | -0.110 | -0.178 | -0.044 | 0.015 | 0.234 | 0.096 |
| | (0.160) | (0.171) | (0.176) | (0.176) | (0.152) | (0.181) |
| Creativity | 0.058 | 0.115 | -0.336 | 0.059 | 0.197 | 0.105 |
| | (0.233) | (0.251) | (0.261) | (0.222) | (0.262) | (0.255) |
| # of misbehaviors during the last week | 0.361 | -0.040 | 0.184 | -0.091 | -0.142 | -0.285 |
| | (0.251) | (0.228) | (0.259) | (0.271) | (0.249) | (0.278) |
| # of nominated friends | -0.179 | -0.025 | -0.059 | -0.034 | -0.127 | -0.046 |
| | (0.329) | (0.288) | (0.315) | (0.116) | (0.128) | (0.130) |
| Perceived current peer relationship quality | -0.052 | 0.059 | -0.038 | 0.085 | 0.018 | -0.042 |
| | (0.063) | (0.045) | (0.059) | (0.057) | (0.049) | (0.058) |
| # of parents' verbal abuse and physical punishment (last semester) | -0.014 | 0.268 | 0.320 | 0.559 | 0.029 | 0.381 |
| | (0.360) | (0.409) | (0.352) | (0.395) | (0.398) | (0.380) |
| Need to express oneself to parents | -0.017 | -0.012 | -0.037 | 0.066 | 0.064 | 0.092 |
| | (0.055) | (0.063) | (0.059) | (0.061) | (0.057) | (0.070) |
| Perceived parental misunderstanding | 0.060 | 0.010 | -0.016 | 0.075 | 0.007 | 0.105 |
| | (0.067) | (0.070) | (0.061) | (0.068) | (0.061) | (0.071) |
| Need for parental listening | 0.056 | 0.042 | -0.090* | -0.033 | 0.074 | 0.083 |
| | (0.057) | (0.057) | (0.053) | (0.065) | (0.062) | (0.063) |
| *Panel C. Parents' parenting style* | | | | | | |

*Continued on next page*

*(Continued) Balance test*

| Variable | AIPC vs. Control | PG vs. Control | High-coverage vs. Low-coverage | Children's needs message vs. No message | Positive effects message vs. No message | Perspective-taking message vs. No message |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Father: | | | | | | |
| Authoritative | -0.097 | -0.047 | 0.180 | -0.126 | 0.008 | -0.162 |
| | (0.188) | (0.154) | (0.171) | (0.170) | (0.187) | (0.170) |
| Authoritarian | 0.072 | -0.033 | 0.066 | 0.201 | 0.046 | 0.305 |
| | (0.143) | (0.155) | (0.158) | (0.167) | (0.157) | (0.185) |
| Permissive | -0.058 | 0.174 | -0.008 | -0.170 | -0.113 | -0.035 |
| | (0.122) | (0.152) | (0.135) | (0.156) | (0.156) | (0.154) |
| Neglectful | 0.013 | 0.059 | -0.077 | 0.011 | 0.007 | 0.249 |
| | (0.097) | (0.102) | (0.104) | (0.131) | (0.125) | (0.150) |
| Mother: | | | | | | |
| Authoritative | 0.021 | 0.083 | 0.101 | -0.088 | -0.176 | -0.070 |
| | (0.197) | (0.158) | (0.187) | (0.172) | (0.153) | (0.178) |
| Authoritarian | -0.010 | -0.299** | -0.030 | 0.158 | -0.117 | 0.156 |
| | (0.164) | (0.150) | (0.156) | (0.154) | (0.165) | (0.181) |
| Permissive | -0.014 | 0.060 | 0.025 | -0.093 | -0.199 | -0.053 |
| | (0.158) | (0.141) | (0.134) | (0.193) | (0.186) | (0.161) |
| Neglectful | 0.042 | -0.004 | -0.028 | -0.041 | -0.080 | 0.027 |
| | (0.076) | (0.075) | (0.075) | (0.099) | (0.116) | (0.107) |
| *Panel D: Child's daily time spent (in minutes)* | | | | | | |
| Doing homework, reviewing, and previewing lessons | -2.379 | -3.271 | -1.539 | 0.505 | -0.375 | 1.822 |
| | (2.243) | (2.301) | (2.212) | (1.881) | (1.708) | (1.812) |
| Extracurricular reading | 0.894 | -0.852 | -2.054 | 0.244 | -1.098 | -0.523 |
| | (1.377) | (1.296) | (1.334) | (1.255) | (1.247) | (1.385) |
| Using Internet and playing video games | 0.541 | -0.905 | 0.597 | -0.285 | -0.516 | 2.069 |
| | (1.618) | (1.339) | (1.608) | (1.429) | (1.336) | (1.409) |
| Watching TV | -1.294 | -1.581 | 1.026 | -0.078 | -2.110 | 0.151 |
| | (1.584) | (1.472) | (1.482) | (1.618) | (1.476) | (1.567) |
| Doing sports/physical activities | 0.359 | -0.185 | 0.119 | -0.811 | 0.021 | 0.802 |
| | (1.463) | (1.508) | (1.375) | (1.613) | (1.452) | (1.701) |
| Playing with friends/classmates | -0.466 | -0.828 | 0.403 | -2.461 | -0.978 | 0.129 |
| | (1.442) | (1.437) | (1.610) | (1.851) | (2.152) | (1.910) |
| Playing alone | -1.378 | -2.380* | 1.741 | -0.438 | -1.929 | 1.430 |
| | (1.631) | (1.423) | (1.262) | (1.675) | (1.410) | (1.867) |

*Continued on next page*

*(Continued) Balance test*

| Variable | AIPC vs. Control | PG vs. Control | High-coverage vs. Low-coverage | Children's needs message vs. No message | Positive effects message vs. No message | Perspective-taking message vs. No message |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Studying with parental supervision | -0.555 | -1.384 | 0.663 | 0.017 | 0.462 | 1.079 |
| | (1.617) | (1.620) | (1.908) | (1.637) | (1.640) | (2.079) |
| Playing with parents | -0.420 | 0.160 | 1.252 | -1.263 | -0.575 | 1.357 |
| | (1.309) | (1.409) | (1.391) | (1.606) | (1.313) | (1.491) |
| Listening to stories told by parents | 0.659 | -0.085 | 0.423 | -1.309 | -1.156 | 1.542 |
| | (1.347) | (1.258) | (1.350) | (1.516) | (1.310) | (1.784) |
| School-grade FE | ✓ | ✓ | | | | |
| Guidance-group FE | | | ✓ | | | |
| Class FE | | | | ✓ | ✓ | ✓ |

*Notes*: This table reports balance tests for variables from non-left-behind children's report. Columns (1)–(2) test for differences between the AI guidance, booklet guidance, and no guidance groups, conditional on school–grade fixed effects. Column (3) tests for differences between the high-coverage and low-coverage groups, conditional on guidance-group fixed effects. Columns (4)–(6) test for differences between the children's needs message group, the positive effects message group, the perspective-taking message group, and the no-message group, conditional on class fixed effects. Chinese and math test scores are standardized within each school-by-grade cell to have a mean of 0 and a standard deviation of 1. Standard errors clustered at the class level are reported in parentheses.

Table 6: Balance test for Variables From Parents' Report (Non-Left-Behind Children)

| Variable | AIPC vs. Control | PG vs. Control | High-coverage vs. Low-coverage | Children's expressed needs vs. No message | Benefits of supportive parenting vs. No message | Perspective-taking reflection vs. No message |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A. Parents' characteristics* | | | | | | |
| Father's age | -0.301 | -0.019 | -0.064 | 0.392 | 0.265 | 0.078 |
| | (0.226) | (0.213) | (0.198) | (0.275) | (0.253) | (0.257) |
| Mother's age | 0.041 | -0.100 | -0.279 | 0.257 | 0.197 | -0.144 |
| | (0.221) | (0.197) | (0.195) | (0.265) | (0.250) | (0.253) |
| Father' years of schooling | -0.072 | -0.026 | -0.305 | -0.129 | -0.251 | -0.017 |
| | (0.170) | (0.148) | (0.282) | (0.169) | (0.184) | (0.194) |
| Mother' years of schooling | 0.121 | 0.114 | -0.039 | -0.208 | -0.298 | -0.146 |
| | (0.190) | (0.165) | (0.286) | (0.178) | (0.194) | (0.201) |
| Father employed in the local county | 0.013 | 0.021 | 0.016 | 0.005 | 0.031 | 0.029 |
| | (0.026) | (0.023) | (0.026) | (0.026) | (0.028) | (0.027) |
| Mother employed in the local county | -0.011 | -0.008 | -0.001 | 0.003 | 0.026* | -0.008 |
| | (0.014) | (0.013) | (0.013) | (0.016) | (0.013) | (0.015) |
| *Panel B. Parents' AI usage and perceived role in parenting* | | | | | | |
| Using AI for child education | -0.003 | -0.010 | -0.008 | 0.012 | -0.008 | -0.010 |
| | (0.013) | (0.011) | (0.011) | (0.015) | (0.014) | (0.017) |
| Using AI for own study/work | -0.010 | 0.003 | -0.027 | -0.027 | -0.008 | 0.005 |
| | (0.019) | (0.021) | (0.025) | (0.028) | (0.026) | (0.028) |
| Using AI for other purposes | -0.008 | -0.014 | -0.006 | -0.016 | -0.013 | 0.007 |
| | (0.013) | (0.013) | (0.014) | (0.020) | (0.021) | (0.020) |
| Perceived role of AI in parenting | 0.018 | -0.164 | 0.089 | 0.295 | 0.163 | 0.294 |
| | (0.249) | (0.274) | (0.253) | (0.336) | (0.312) | (0.284) |
| *Panel C. Parents' evaluation of their children* | | | | | | |
| Satisfaction with academic performance in the last semester | -0.036 | -0.081 | 0.048 | 0.013 | -0.014 | 0.128 |
| | (0.103) | (0.103) | (0.112) | (0.118) | (0.100) | (0.105) |
| Rating of academic performance in the last semester | 0.004 | -0.736 | 0.043 | -0.742 | -1.501 | -0.904 |
| | (1.012) | (1.026) | (1.031) | (1.069) | (1.140) | (1.137) |
| Expectation for the child's academic achievement | 0.003 | 0.027 | -0.009 | -0.034 | -0.026 | -0.023 |
| | (0.029) | (0.026) | (0.027) | (0.028) | (0.031) | (0.027) |
| *Panel D. Parents' outcomes and their parenting practices* | | | | | | |

*Continued on next page*

*(Continued) Balance test*

| Variable | AIPC vs. Control | PG vs. Control | High-coverage vs. Low-coverage | Children's needs message vs. No message | Positive effects message vs. No message | Perspective-taking message vs. No message |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Mental health | -0.049 | 0.304 | -0.054 | -0.138 | -0.018 | 0.014 |
| | (0.275) | (0.264) | (0.253) | (0.244) | (0.251) | (0.247) |
| # of parent–child conflicts during the last semester | 1.059 | 0.839 | -0.845 | 0.375 | 0.812 | 0.420 |
| | (0.864) | (0.765) | (0.660) | (0.948) | (0.855) | (0.740) |
| *Parenting skills in the last semester:* | | | | | | |
| Rule enforcement | 0.013 | 0.013 | 0.057 | 0.017 | -0.036 | 0.041 |
| | (0.078) | (0.076) | (0.074) | (0.090) | (0.077) | (0.075) |
| Emotional regulation | -0.111 | -0.073 | -0.005 | 0.064 | 0.053 | 0.130 |
| | (0.078) | (0.069) | (0.067) | (0.079) | (0.085) | (0.078) |
| Communication skills | 0.115 | 0.089 | 0.030 | -0.172 | -0.088 | 0.028 |
| | (0.097) | (0.109) | (0.096) | (0.104) | (0.104) | (0.094) |
| Positive parenting practices | 0.049 | 0.026 | 0.153 | 0.005 | 0.136 | 0.010 |
| | (0.174) | (0.169) | (0.176) | (0.159) | (0.157) | (0.167) |
| Parenting self-efficacy | -0.043 | -0.143 | 0.305 | 0.165 | 0.027 | 0.166 |
| | (0.237) | (0.220) | (0.225) | (0.283) | (0.251) | (0.208) |
| Parenting awareness | 0.149 | -0.045 | -0.047 | -0.063 | -0.868* | -0.221 |
| | (0.458) | (0.445) | (0.482) | (0.435) | (0.467) | (0.405) |
| Punitive discipline measures | 0.116 | 0.066 | -0.056 | 0.067 | -0.017 | -0.006 |
| | (0.075) | (0.070) | (0.069) | (0.097) | (0.079) | (0.093) |
| Command-style discipline measures | 0.049 | 0.069 | -0.099 | -0.092 | -0.128 | -0.133 |
| | (0.100) | (0.100) | (0.085) | (0.098) | (0.101) | (0.094) |
| Permissive discipline measures | 0.090 | 0.018 | 0.010 | -0.010 | -0.057 | 0.036 |
| | (0.055) | (0.051) | (0.053) | (0.049) | (0.041) | (0.045) |
| Communicative discipline measures | -0.258 | -0.160 | 0.155 | 0.100 | 0.223 | 0.141 |
| | (0.174) | (0.159) | (0.156) | (0.143) | (0.135) | (0.142) |
| Empathy | 0.175 | 0.167 | -0.106 | 0.006 | -0.033 | 0.000 |
| | (0.177) | (0.178) | (0.201) | (0.157) | (0.176) | (0.149) |
| *Panel E. Monthly expenditures and household income* | | | | | | |
| Monthly spending on children's school supplies | 5.396 | -4.121 | 3.197 | 2.929 | -4.512 | -3.616 |
| | (3.563) | (3.876) | (3.641) | (4.233) | (3.995) | (3.806) |
| Monthly spending on extracurricular resources | 2.797 | 1.231 | -0.920 | -3.260 | 3.011 | -3.305 |
| | (3.546) | (4.260) | (3.847) | (3.981) | (4.609) | (3.726) |
| Monthly spending on toys | 2.053 | -1.318 | -1.586 | 2.909 | -6.336 | -2.653 |

*Continued on next page*

*(Continued) Balance test*

| Variable | AIPC vs. Control | PG vs. Control | High-coverage vs. Low-coverage | Children's needs message vs. No message | Positive effects message vs. No message | Perspective-taking message vs. No message |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | (4.152) | (5.023) | (4.435) | (4.749) | (4.295) | (4.000) |
| Monthly household income | 130.058 | -306.645 | -258.103 | 84.215 | -82.361 | 227.872 |
| | (199.268) | (204.296) | (198.653) | (218.921) | (244.133) | (220.055) |
| *Panel F. Paternal and maternal grandparents' parenting styles during the parents' childhoods* | | | | | | |
| Paternal grandparents: | | | | | | |
| Authoritative | -0.065 | -0.083 | 0.091 | 0.142* | 0.093 | 0.062 |
| | (0.080) | (0.086) | (0.075) | (0.081) | (0.086) | (0.083) |
| Authoritarian | 0.051 | 0.067 | -0.067 | 0.013 | 0.095 | -0.006 |
| Maternal grandparents: | | | | | | |
| Authoritative | -0.086 | -0.140* | 0.006 | 0.073 | 0.045 | -0.002 |
| | (0.073) | (0.080) | (0.083) | (0.106) | (0.097) | (0.085) |
| Authoritarian | -0.035 | 0.054 | -0.001 | 0.093 | 0.066 | -0.096 |
| | (0.066) | (0.066) | (0.058) | (0.065) | (0.066) | (0.077) |
| School-grade FE | ✓ | ✓ | | | | |
| Guidance-group FE | | | ✓ | | | |
| Class FE | | | | ✓ | ✓ | ✓ |

*Notes*: This table reports balance tests for variables from parents' report, restricting the sample to households in which the child lives with both parents. Columns (1)–(2) test for differences between the AI guidance, booklet guidance, and no guidance groups, conditional on school–grade fixed effects. Column (3) tests for differences between the high-coverage and low-coverage groups, conditional on guidance-group fixed effects. Columns (4)–(6) test for differences between the children's needs message group, the positive effects message group, the perspective-taking message group, and the no-message group, conditional on class fixed effects. Standard errors clustered at the class level are reported in parentheses.

**ONLINE APPENDIX**

**Supplementary material to *Pre-Analysis Plan: Can AI Enhance Parenting Quality for Rural Parents?***

Pengfei Chu, Xueheng Li, Jia Wu, Yiming Xia

# A Theory of Change

**Need.** The intervention addresses two pressing challenges in rural and low-income settings. Parents often lack structured, evidence-based knowledge of child development and effective parenting strategies, which leads to inconsistent discipline, weak communication, and limited emotional support. At the same time, digital tools such as AI advisors have the potential to provide personalized, timely guidance, but parents frequently lack the training or confidence to use them effectively. Traditional booklets raise awareness but rarely sustain behavioral change, while digital tools remain underutilized. This context highlights the need for interventions that combine simple, accessible knowledge dissemination with interactive support.

**Inputs.**

- **T1 (Parenting Booklet):** Parents receive a handbook of parenting skills, which covers evidence-based strategies, child development knowledge, discipline, and everyday routines.
- **T2 (AI-Assisted Parenting):** Parents receive the handbook plus access to an AI parenting advisor. They also receive coaching on how to use the tool and periodic nudge messages to sustain engagement and improve fidelity of use.
- **Reminder nudge**: Parents receive a bimonthly reminder nudge.

**Outputs.**

- **T1:** Parents are expected to read and learn from the handbook, gaining awareness of effective strategies and beginning to apply them in daily parenting.
- **T2:** Parents are expected to engage actively with the AI tool, receive personalized and context-specific guidance, and incorporate these strategies more consistently into their parenting practices, supported by coaching and reminders.
- **Reminder nudge**: Parents are expected to engage more actively with the AI tool or handbook.

**Intermediate Outcomes.**

- **T1:** Parents improve their knowledge and skills, including understanding child development and effective parenting strategies, discipline, and daily routines. They also demonstrate improvements in parenting behaviors: higher quality parent–child interactions, more consistent discipline, fewer harsh or impulsive reactions, greater warmth and responsiveness, more effective communication, and stronger conflict resolution skills.
- **T2:** Parents achieve the same improvements as in T1, with the additional development of skills in using AI tools to support parenting. The personalized and interactive nature of the AI guidance is expected to strengthen adoption and sustain positive changes in behavior.
- **Reminder nudge**: Parents are prompted to develop parenting skills and engage more with the AI tool.

**Final Outcomes.**

- **Parents:** Both T1 and T2 are expected to improve parental emotional regulation and psychological well-being, with reduced stress and anxiety, greater calmness, and stronger self-efficacy in parenting. These improvements are expected to be larger under T2. The reminder nudge is expected to generate additional gains.
- **Children:** Children benefit from improvements in parenting practices and home environments. Expected outcomes include fewer behavioral problems, greater cooperation, stronger prosocial behavior, enhanced socioemotional development, and improved academic and cognitive performance.

**Impact.** By comparing a benchmark approach with an AI-assisted approach, the program generates evidence on whether personalized digital support adds value beyond traditional knowledge dissemination. This comparison aims to identify scalable and cost-effective strategies to strengthen parenting practices and promote children's development in rural settings.

# B Details about information intervention

## B.1 Measurement of children's needs

We use the following survey items to capture whether children want greater understanding and respect from their parents.

*Please answer the following questions based on your experiences last semester:*

1. *I want my parents to listen more to what I think.*
2. *I want my parents to give me more chances to express myself.*
3. *I feel that my parents don't understand me.*

These items capture, respectively, children's desire for parental understanding and respect, for greater opportunities to express themselves, and for more attentive listening from parents. Response options are *not at all*, *rarely*, *sometimes*, and *often*. We code each need as present if the child responded *sometimes* or *often*. We then compute, for each school, the share of students reporting each need. Descriptive statistics for these variables are reported at the bottom of Panel B in Table 2.

## B.2 Estimating the effects of patient and respectful parenting

### B.2.1 Measurement of key variables

**Patient and respectful parenting.** We measure the extent to which parents treat their child patiently and respectfully using four survey items reported by the child.

**Parental responses to children's mistakes.** *Based on the previous semester, how many times did the following occur between you and your parents?*

1. *When I made a mistake, my father did not give me a chance.*
2. *When I made a mistake, my mother did not give me a chance.*

Response options are 0, 1, 2, 3, 4, 5, 6, and 7 or more times. We code these responses as 0–7, with 7 indicating "7 or more." Because these items are negatively worded, we reverse-code them so that higher values indicate more patient and respectful parenting.

**Parental explanations for following rules.** We also use two items to capture whether parents explain why rules should be followed:

1. *My father explains why I should follow rules.*
2. *My mother explains why I should follow rules.*

Response options are *strongly disagree*, *disagree*, *neutral*, *agree*, and *strongly agree*. We code responses from 1 to 5, where 1 = *strongly disagree* and 5 = *strongly agree*.

Within each school, we combine the four items into a single score using principal component analysis (PCA) and standardize the resulting score to have a mean of zero and a standard deviation of one. We then construct a binary indicator for patient and respectful parenting, coded as 1 if the standardized score is above 0 and 0 otherwise.

**Child anxiety.** We measure children's anxiety using the following item:

> *Do you currently experience the following?*
> *I feel anxious and easily unsettled.*

Response options are *strongly disagree*, *disagree*, *somewhat disagree*, *neutral*, *somewhat agree*, *agree*, and *strongly agree*. We further code responses on a seven-point Likert scale from 1 to 7, where *strongly disagree*=1, *disagree*=2, *somewhat disagree*=3, *neutral*=4, *somewhat agree*=5, *agree*=6, and *strongly agree*=7.

**Independent thinking.** We measure whether children perceive themselves as thinking independently using the following item:

> *Please answer the following question based on your current situation:*
> *I like doing things the same way and do not like trying new ways.*

Response options are *not at all true*, *not true*, *somewhat true*, and *completely true*. We reverse-code this item so that higher values indicate stronger independent thinking. Specifically, responses are coded from 1 to 4, where *not at all true*=4, *not true*=3, *somewhat true*=2, and *completely true*=1.

### B.2.2 Effect estimation

Within each school, we estimate separate regressions of the outcomes—child anxiety and independent thinking—on the indicator for patient and respectful parenting. To facilitate interpretation across outcomes and schools, we scale each regression coefficient by the within-school mean of the corresponding dependent variable; thus, the reported estimates can be interpreted as the proportional (percent) change in the outcome associated with patient and respectful parenting. The resulting coefficients provide school-specific estimates of the effect of patient and respectful parenting.

# C Questionnaires

## C.1 Children surveys

### C.1.1 Delay of gratification

Children's willingness to delay gratification is measured using a simple hypothetical intertemporal choice task. Each child is presented with the following question:

> Based on the following two scenarios A and B, please make your choice.
> *I choose* _____.

The options are:

- Option A: Receive 5 yuan today.
- Option B: Receive 10 yuan one week from now.

We construct a binary variable, *Delay of gratification*, which equals 1 if the child chooses the delayed payment (option B) and 0 if the child chooses the immediate payment (option A).

### C.1.2 Risk preference

Children's risk-taking preference is measured using a simple hypothetical risky choice task. Each child is presented with the following question:

> Based on the following two scenarios A and B, please make your choice.
> *I choose* _____.

The options are:

- Option A: Receive 5 yuan for sure.
- Option B: Toss a coin: if heads, receive 10 yuan; if tails, receive 0 yuan.

We construct a binary variable, *Risk preference*, which equals 1 if the child chooses the risky lottery (option B) and 0 if the child chooses the safe option (option A). Higher values therefore indicate a greater willingness to take risk.

### C.1.3 Frequency of parents' verbal abuse and physical punishment during the last semester

Frequency of parents' verbal abuse and physical punishment is measured with a four–item checklist. The instructions read:

> Regarding your relationship with your parents, please think about the last semester and choose how many times each of the following has happened.

The items are:

1. Your father scolds you.
2. Your father hits you.
3. Your mother scolds you.
4. Your mother hits you.

For each item, the child chooses one of eight frequency categories: 0, 1, 2, 3, 4, 5, 6, or 7 or more times during the current school term. Frequency of parents' verbal abuse and physical punishment is the sum of these codes across all four items; higher values indicate more frequent and severe abuse.

### C.1.4 Mental health

Children's mental health condition is measured using the CES-D 10 scale. For each of the ten statements below, children are asked to think about their feelings during the past week and choose one of four options: *rarely (0 days)*, *sometimes (1–2 days)*, *often (3–4 days)*, or *most of the time (5–7 days)*. These options are coded from 0 to 3, respectively.

The items are:

1. I am often bothered by small things.
2. I find it hard to concentrate on what I am doing.
3. I feel down or depressed.
4. I feel that doing anything takes a lot of effort.
5. I feel hopeful about the future.
6. I am often worried or fearful about some things.
7. My sleep is not good.
8. I often feel happy.
9. I feel lonely.
10. I feel that life is meaningless.

Among these items, two are positively worded and require reverse coding:

- I feel hopeful about the future.
- I often feel happy.

These two items are reverse-coded so that higher values consistently indicate worse mental health (more depressive symptoms).

After reverse coding the two positive items and retaining the original coding for the negative items, the CES-D 10 total score is obtained by summing the ten item scores, yielding a range from 0 to 30. A higher total score reflects more severe depressive symptoms.

### C.1.5 Self-control

Children's self-control is measured with a multi-item scale. Children are asked to think about their *current* situation and, for each statement, choose the option that best describes them. The response options are: "strongly disagree", "disagree", "neutral", "agree", and "strongly agree", coded from 1 to 5 respectively (with higher values indicating higher self-control after recoding, as described below).

The items are:

- I do not make plans in advance.
- My time management is poor.
- I am an impulsive person.
- I do not think carefully before I do things.

Negatively worded items that reflect low self-control are reverse-coded so that higher values always indicate better self-control. We then sum the recoded items to construct a self-control index, where higher scores correspond to higher levels of self-control.

### C.1.6 Prosocial behavior, internalizing and externalizing behavior problems

These variables are measured using items from the Strengths and Difficulties Questionnaire (SDQ). Children are asked to think about their *current* experiences and, for each statement, indicate whether it is *not true*, *somewhat true*, or *certainly true*. These three options are coded 1, 2, and 3, respectively (with reverse-coding for some items as described below).

The ten SDQ items are:

1. I am often kind to younger children.
2. If someone is hurt, upset, or feeling ill, I usually offer to help.
3. I am often worried about many things.
4. I am often unhappy, downhearted, or tearful.
5. I get very angry and often lose my temper.
6. I am sometimes rude or mean to other children.
7. I am easily distracted; I find it difficult to concentrate.
8. I usually act without thinking things through.
9. I generally play alone or keep to myself.
10. Other children or young people pick on me or bully me.

**Prosocial behavior.** Prosocial behavior is based on items 1 and 2. Higher values indicate more prosocial behavior. We sum the scores of these two items to construct a prosocial behavior scale.

**Internalizing behavior problems.** Internalizing behavior problems are based on items 3 and 4 (worry and sadness) and item 9 and 10 (social withdrawal). We code each item from 1 to 3 as described above and sum them to obtain an internalizing problems index, where higher scores indicate more severe internalizing problems.

**Externalizing behavior problems.** Externalizing behavior problems are based on items 5, 6, 7, and 8, which capture anger, conduct problems, hyperactivity, and peer aggression/victimization. Each item is coded from 1 to 3 and then summed to obtain an externalizing problems index, with higher scores indicating more severe externalizing behavior problems.

### C.1.7 Altruism

Altruism is measured with a simple dictator game. Each child is told:

> You have 8 candies. You can choose to keep them for yourself or give some of them to another classmate. How many candies would you give to the other child? _____

Children are allowed to choose any integer amount from 0 to 8. In the questionnaire, the options are shown as: 0 candies, 1 candy, 2 candies, 3 candies, 4 candies, 5 candies, 6 candies, 7 candies, 8 candies.

We construct an *altruism* variable equal to the number of candies the child chooses to give away (ranging from 0 to 8). Higher values indicate a greater level of altruism.

### C.1.8 Big Five personality traits

We measure children's personality using a short Big Five inventory adapted from Gosling et al. (2003). Children are asked to think about their *current* situation and, for each statement, choose one of seven options: "strongly disagree", "disagree", "somewhat disagree", "neutral", "somewhat agree", "agree", or "strongly agree". These options are coded from 1 to 7, respectively.

The ten items are:

1. I feel that I am outgoing and enthusiastic.
2. I feel that I like challenges and competition.
3. I feel that I am reliable and self-disciplined.
4. I feel that I am anxious and easily worried.
5. I feel that I enjoy new experiences and trying new things.
6. I feel that I am reserved and calm.
7. I feel that I am empathetic and warm-hearted.
8. I feel that I am careless and lack consideration for others.
9. I feel that I am calm and emotionally stable.
10. I feel that I am conservative and lack creativity.

Items that are negatively worded with respect to desirable personality traits (items 2, 4, 6, 8, 10) are reverse-coded so that higher scores always indicate more positive personality characteristics. We then aggregate the items into five dimension-specific indices, corresponding to the Big Five personality traits: *Extraversion*, *Agreeableness*, *Conscientiousness*, *Neuroticism*, and *Openness*. For each trait, we take the (reverse-coded) items that conceptually belong to that dimension and compute the average score to obtain a trait-specific index. Higher values of each index indicate a higher level of the corresponding personality trait.

### C.1.9 Growth mindset

We measure children's growth mindset using a short scale. Children are asked to read each sentence and, based on their *current* situation, choose one of five options: "strongly disagree", "disagree", "neutral", "agree", or "strongly agree". These options are coded from 1 to 5, respectively.

The three items are:

1. My intelligence is fixed, and I can hardly change it.
2. Intelligence is a quality I was born with, and I cannot really change it.
3. I can learn new knowledge, but I cannot truly change my intelligence.

These items are negatively worded with respect to a growth mindset (they reflect a fixed mindset). We therefore reverse-code these items so that higher scores always indicate a stronger growth mindset. After reverse-coding, we sum all three items to construct a continuous growth mindset index, where higher values correspond to a stronger belief that abilities can be improved through effort.

### C.1.10 Creativity

We measure children's creativity using a nine-item scale. Children are asked to think about their *current* situation and, for each statement, indicate how well it describes them. The response options are: "strongly disagree", "disagree", "agree somewhat", and "strongly agree", coded from 1 to 4, respectively.

The nine items are:

1. I like to do things in the same way and do not like to look for new ways of doing things.
2. I think that every problem has a fixed answer.
3. When I grow up, I want to do something that no one else has thought of.
4. I do not like to be restricted by too many rules.
5. Trying new games and activities is something interesting.
6. I like to think of some new ideas, even if these ideas are not very useful.
7. I like things that are different from what most people like.
8. For many things I really want to have a try.
9. I always hope to do things in the way my parents expect.

Items that reflect lower creativity or conformity (items 1, 2, and 9) are reverse-coded so that, for all items, higher scores indicate higher creativity. We then sum the nine recoded items to construct a creativity index, where higher values represent a higher level of creativity.

### C.1.11 Frequency of misbehaviors during the last semester

We measure students' misbehaviors during the previous semester using a checklist of ten behaviors. For each item, students report how often it occurred, choosing one of four options: "never", "rarely", "sometimes", or "often". These options are coded from 1 to 4, respectively, with higher values indicating more frequent misbehavior.

The ten behaviors are:

1. Exam performance was unsatisfactory and did not meet parents' expectations.
2. Homework was not completed on time or not completed with sufficient quality.
3. Irregular sleep schedule, often staying up late or getting up very late.
4. Bedroom or desk was very messy and unhygienic.
5. Went out to play after school and did not return home immediately.
6. Did not complete homework independently.
7. Told lies.
8. Did not study hard (e.g., did not preview lessons or complete summer homework).
9. Talked back to or contradicted others.
10. Quarreled with parents because of excessive time spent playing on a mobile phone.

We construct the *frequency of misbehaviors* index by summing the coded responses to all ten items. The resulting score ranges from 10 to 40, with higher scores indicating more frequent or more numerous misbehaviors.

### C.1.12 Number of nominated friends

To measure children's friendship networks, we ask each student to list, from their *current* classmates, up to eight peers with whom they have the best relationships. The item is introduced as:

> Please write down the names of up to 8 classmates in your current class who are your best friends (if you have fewer than 8 friends, you may leave the remaining spaces blank).

We then construct a variable *number of nominated friends*, defined as the total count of friends a student lists (ranging from 0 to 8). A higher value indicates that the student nominates more friends within the class.

### C.1.13 Perceived current peer relationship quality

We measure negative aspects of peer relationships using four items about classmates' behaviors toward the respondent. Students are asked, based on their *current* situation, to indicate for each statement whether such experiences occurred: the options are "yes" and "no", coded as 0 and 1, respectively.

The items are:

1. Classmates in my class often insult or verbally abuse me.
2. Classmates in my class often hit or push me.

3. I am often excluded by classmates in my class.
4. Classmates in my class often laugh at or mock me.

We construct a *peer relationship* index by summing the four binary items (range 0–4), where higher values indicate great peer relationship quality (i.e., less frequent experiences of bullying or peer victimization).

### C.1.14 Parents' parenting styles

Following Maccoby and Martin (1983), we measure parenting styles separately for fathers and mothers using the 11-item scale. Children are asked to describe, based on their *last semester's* experience, how well each statement matches their father's (and then their mother's) typical behavior. For each item, they choose one of five options: "strongly disagree", "disagree", "neutral", "agree", or "strongly agree", coded from 1 to 5, respectively.

The 11 items (asked once about the father and once about the mother) are:

1. Even when we have different opinions, my father (mother) encourages me to express my own ideas.
2. My father (mother) explains to me why I should obey the rules.
3. When I am in difficulty, my father (mother) comforts and understands me.
4. My father (mother) uses hitting as a way to discipline me.
5. When I ask why I must obey, my father (mother) answers "because I said so".
6. My father (mother) uses scolding or criticism to push me to make progress.
7. When I am in a bad mood, my father (mother) gives in to me.
8. My father (mother) spoils me.
9. My father (mother) announces a punishment, but in fact does not carry it out.
10. My father (mother) rarely asks about how I am doing at school.
11. When I feel distressed, my father (mother) usually does not come to comfort me.

We use different subsets of items to construct four parenting-style indices for fathers and mothers separately:

- **Authoritative** (items 1–3): behaviors combining warmth and reasonable guidance.
- **Authoritarian** (items 4–6): harsh or coercive discipline.
- **Permissive** (items 7–9): indulgent or inconsistent discipline.
- **Neglectful** (items 10–11): low involvement or emotional neglect.

For each parent and each style, we sum the corresponding item scores to obtain four continuous indices for the father and four for the mother. Higher values indicate that the given parenting style is more strongly perceived by the child.

### C.1.15 Children's daily time use

We measure children's daily time use using children's *self-reports*. In the questionnaire, each child was asked to report, for the last week, how much time they spent at home in each of the following activities, separately for weekdays and weekends:

1. Time spent doing homework, reviewing, and previewing lessons.
2. Time spent on extracurricular reading.
3. Time spent using the Internet or playing computer/video games.
4. Time spent watching TV.
5. Time spent in sports/physical activities.
6. Time spent playing with friends/classmates.
7. Time spent playing alone.
8. Time spent studying with parents.
9. Time spent playing with parents.
10. Time spent with parents telling stories, chatting, and discussing issues.

For each activity, children reported the *average number of minutes per day* from Monday to Friday and from Saturday to Sunday, choosing one lettered category:

(A) 0 minutes; (B) 1–10; (C) 11–20; (D) 21–30; (E) 31–40; (F) 41–50; (G) 51–60; (H) 61–70; (I) 71–80; (J) 81–90; (K) 91–120; (L) more than 120 minutes.

In the data, these options are first coded as integers: A=0, B=1, C=2, ..., L=11. We then map these codes to minutes using the midpoint of each interval (0 for A, 5 for B, 15 for C, 25 for D, 35 for E, 45 for F, 55 for G, 65 for H, 75 for I, 85 for J, 105 for K, and 120 for L).

For each activity $j$, we compute the child's *average daily time* as the average of weekdays and weekends:

$$\text{Time use}_j = \frac{T_{j,\text{Mon–Fri}} + T_{j,\text{Sat–Sun}}}{2},$$

where $T_{j,\text{Mon–Fri}}$ and $T_{j,\text{Sat–Sun}}$ are the minute values converted from the categorical responses for Monday–Friday and Saturday–Sunday, respectively. Thus, each constructed variable reflects the child-reported average number of minutes per day spent on that activity; higher values indicate more time spent.

## C.2 Parents surveys

### C.2.1 Parents' evaluation and expectations of the child's academic performance

The parent questionnaire elicits parents' assessments and expectations regarding the child's academic performance through three items.

**Satisfaction with the child's academic performance.** Parents report how satisfied they are with the child's academic performance in the last semester on a 10-point scale, where 1 indicates "very dissatisfied" and 10 indicates "very satisfied". We use this raw score as a continuous variable; higher values indicate greater parental satisfaction.

**Rating of the child's academic performance.** Parents are asked: if you were to give your child a score for his/her academic performance in the last semester (full score = 100), what score would you give? The answer is an open-ended number between 0 and 100. We treat this as a continuous measure of the parent's subjective rating of the child's academic performance.

**Expectation for the child's academic achievement.** Parents are asked how high their expectations are for the child to attain good academic results, with three options: (A) very high expectations, (B) average expectations, and (C) no expectations.

In our dataset, this item is reverse-coded to obtain a variable ranging from 1 to 3:

$$1 = \text{no expectations}; \quad 2 = \text{average expectations}; \quad 3 = \text{very high expectations}.$$

Thus, higher values indicate *higher* parental expectations for the child's academic achievement.

### C.2.2 Perceived role of AI in parenting

We measure parents' attitudes toward the role of AI tools in parenting with four items. Parents are asked: "What do you think is the role of AI tools in educating your child?" For each statement, they indicate their level of agreement using a 7-point Likert scale: *strongly disagree*, *disagree*, *somewhat disagree*, *neutral*, *somewhat agree*, *agree*, or *strongly agree*. These options are coded from 1 to 7, respectively, so that larger values indicate a more positive view of AI in parenting.

The items are:

- AI can improve parent–child communication methods.
- AI can help me understand my child's emotions and behaviors.
- AI can provide useful educational suggestions.
- AI can enhance the quality of the parent–child relationship.

We sum the four item scores to construct an index of *perceived role of AI in parenting*, ranging from 4 to 28. Higher scores reflect stronger beliefs that AI tools play a beneficial role in communication, understanding the child, offering educational advice, and improving parent–child relationship quality.

### C.2.3 Mental health

Parents' mental health condition is measured using a CES-D 10 scale. For each of the ten statements below, parents are asked to think about their feelings during the past week and choose one of four options: *rarely (0 days)*, *sometimes (1–2 days)*, *often (3–4 days)*, or *most of the time (5–7 days)*. These options are coded from 0 to 3, respectively.

The items are:

1. I am often bothered by small things.
2. I find it hard to concentrate on what I am doing.
3. I feel down or depressed.
4. I feel that doing anything takes a lot of effort.
5. I feel hopeful about the future.
6. I am often worried or fearful about some things.
7. My sleep is not good.
8. I often feel happy.
9. I feel lonely.
10. I feel that life is meaningless.

Among these items, two are positively worded and require reverse coding:

- I feel hopeful about the future.
- I often feel happy.

These two items are reverse-coded so that higher values consistently indicate more severe mental distress (more depressive symptoms).

After reverse coding the two positive items and retaining the original coding for the negative items, the CES-D 10 total score is obtained by summing the ten item scores, yielding a range from 0 to 30. A higher total score reflects more severe depressive symptoms.

### C.2.4 Parent-child conflicts during the last semester

We measure the frequency of conflicts between parents and children during the last semester using parents' self-reports. Parents are asked: "Based on the situation in the last semester, how often did the following types of conflicts occur between you and your child?" For each item, they choose one of eight options: *0, 1, 2, 3, 4, 5, 6, or 7 or more times*. In the data, these options are coded from 0 to 7, respectively (with "7 or more times" coded as 7).

The ten conflict situations are:

- The child's test scores are unsatisfactory and fall short of expectations.
- The child delays and fails to complete homework on time.
- The child's homework quality is poor and contains many mistakes.
- The child overuses electronic products or is addicted to games / short videos.
- The child is picky about food or wastes food.
- The child keeps irregular hours, staying up late or getting up late.
- The child's room or desk is messy, and the child ignores hygiene.
- The child spends money without restraint.

- The child performs poorly at school and is criticized by teachers.
- The child refuses to attend after-school tutoring classes.

For each item, a higher value indicates that this type of conflict occurred more frequently. e construct an overall conflict index by summing the ten item scores, where higher values represent more frequent parent–child conflicts over the semester.

### C.2.5 Parenting skills

We measure parenting skills using parents' self-reports on six hypothetical scenarios about their interactions with the child during the last semester. For each scenario, parents are asked which of five responses best describes what they would usually do (options A–E). In the data, these options are coded from 1 to 5, with higher values indicating a less desirable response (e.g., more inconsistent rule enforcement, poorer emotional control, or less effective communication).

The six items are:

1. If your child procrastinates and does not complete homework, how do you usually respond?
2. If your child violates family rules, how do you usually respond?
3. If your child's exam results are unsatisfactory, how do you usually respond emotionally (e.g., staying calm vs. losing temper)?
4. If your child provokes you, how well can you control your emotions?
5. When your child does something wrong, how clearly and patiently do you explain the mistake and its consequences?
6. If your child does not follow the agreed schedule, how do you usually remind them (e.g., direct, brief reminder vs. repeated nagging)?

Following the questionnaire design, we group these items into three dimensions of parenting skills:

- **Rule enforcement** (items 1 and 2): consistency and firmness in requiring the child to follow rules and complete assigned tasks.
- **Emotional regulation** (items 3 and 4): ability to remain calm and regulate emotions when facing the child's poor performance or provocation.
- **Communication skills** (items 5 and 6): clarity, brevity, and effectiveness of communication when correcting the child or reminding them about routines.

For each dimension, we sum the corresponding item scores to construct three indices: *rule enforcement*, *emotional regulation*, and *communication skills*. All indices are coded so that higher values indicate better parenting skills in the corresponding dimension.

### C.2.6 Positive parenting practices

We assess positive parenting practices using four items from the parent questionnaire. Parents are asked to indicate, based on their experience in the last semester, how well each statement describes their typical behavior with the child. For every item, they choose one of five options: *completely untrue*, *somewhat untrue*, *neutral*, *somewhat true*, or *completely true*. These options are coded from 1 to 5, respectively, so that larger values indicate more positive parenting.

The items are:

- When my child disagrees with me, I encourage him/her to express his/her own ideas.
- When my child violates family rules, I remain calm and handle the situation in a reasonable way.
- I encourage my child to take age-appropriate responsibilities (such as completing homework or tidying up belongings).
- When my child feels down, I comfort him/her and help him/her regulate emotions.

We construct a *positive parenting practices* index by summing the four item scores. The index ranges from 4 to 20, with higher values indicating that the parent more frequently engages in supportive, respectful, and emotionally responsive parenting behaviors.

### C.2.7 Parenting self-efficacy

Parents' confidence in their parenting ability is measured using six items on parenting self-efficacy. Parents are asked to read several statements about parenting and, based on their actual situation, choose the option that best describes them during the last semester. For each item, they choose one of five options: *strongly disagree*, *disagree*, *neutral*, *agree*, or *strongly agree*. These options are coded from 1 to 5, respectively.

The items are:

- I feel that I can handle my child's problems well.
- When my child does not listen, I know how to make him/her cooperate.
- I can find ways to solve difficulties in raising my child.
- Being a father/mother makes me feel happy.
- I very much enjoy the role of being a father/mother.
- I feel that taking care of my child is worthwhile.

We sum the six item scores to obtain a *parenting self-efficacy* scale ranging from 6 to 30, where higher values indicate stronger perceived competence and satisfaction in the parenting role.

### C.2.8 Parenting awareness

Parenting awareness is measured with eight items capturing parents' beliefs about effective parenting practices and the consequences of different disciplinary styles. Parents are asked: "Do you agree with the following statements?" and, based on their actual views, select one of five options: *strongly disagree*, *disagree*, *neutral*, *agree*, or *strongly agree*. These options are coded from 1 to 5, respectively, so that larger values indicate stronger endorsement of these awareness statements.

The items are:

- Frequent scolding and punishment will undermine a child's intrinsic motivation.
- Children who are often scolded tend to avoid problems and show less trust in their parents.
- Command-style education easily triggers psychological reactance and harms children's concentration and self-confidence.
- Only giving orders without teaching methods is not conducive to improving a child's task-management ability.
- Long-term indulgence weakens a child's time-management skills and habits.
- Long-term indulgence is not conducive to cultivating children's sense of rules and responsibility.
- Over-satisfying a child's demands will make the child lack frustration tolerance.
- Excessive parental pampering or indulgence undermines a child's independence and self-management ability.

We sum the eight item scores to construct a *parenting awareness* scale ranging from 8 to 40, where higher scores reflect greater awareness of the potential negative effects of harsh discipline and overindulgence, and stronger endorsement of more developmentally appropriate parenting concepts.

### C.2.9 Parenting discipline practices

We measure parents' discipline practices in response to common child behaviors using nine vignettes. Parents are asked:

> "When your child shows the following behaviors, which type of educational practice do you usually adopt during the last semester ?"

For each vignette, parents choose one of four responses: *punitive*, *command-style*, *laissez-faire*, or *communicative*. The nine situations are:

1. The child's exam performance is unsatisfactory and below expectations.
2. The child fails to complete homework on time, or the homework quality is poor.
3. The child overuses electronic devices and is addicted to games or short videos.
4. The child is picky about food or wastes food.
5. The child has irregular routines, such as staying up late or getting up late.
6. The child keeps the room or desk messy and does not pay attention to hygiene.
7. The child spends money without restraint.
8. The child performs poorly at school and is criticized by the teacher.
9. The child refuses to attend after-school tutoring classes.

The four response options are defined as follows:

- **Punitive**: using punishment to make the child obey.

- **Command-style**: requiring the child to comply immediately in a commanding tone without discussion or explanation.
- **Laissez-faire**: neither reminding nor setting limits, letting the child handle it on their own.
- **Communicative**: discussing the reasons with the child, jointly agreeing on a solution, and then reminding or checking according to the agreement.

Based on parents' choices across the nine scenarios, we construct four variables:

- **Punitive discipline frequency**
- **Command-style discipline frequency**
- **Laissez-faire discipline frequency**
- **Communicative discipline frequency**

Each variable equals the number of scenarios (0–9) in which the parent selects the corresponding option; higher values indicate more frequent use of that discipline practice.

### C.2.10 Empathy

Parents' empathy is assessed with four items. Parents are asked to read several statements about empathic ability and, based on their actual situation, choose the option that best describes them. For each item, they choose one of five options: *completely untrue*, *somewhat untrue*, *neutral*, *somewhat true*, or *completely true*. These options are coded from 1 to 5, respectively.

The items are:

- I often put myself in others' shoes to understand their thoughts.
- When others disagree with me, I try to look at the issue from their perspective.
- When I see someone in difficulty, I feel concerned for them.
- When others feel sad, I am emotionally affected by their feelings.

We sum the four item scores to construct an *empathy* scale ranging from 4 to 20, where higher scores indicate stronger empathic ability.

### C.2.11 Paternal grandparents' parenting styles during the parents' childhood

We measure paternal grandparents' parenting styles based on the father's retrospective reports about how his parents (i.e., the child's paternal grandparents) raised him during his childhood. Fathers are asked to evaluate the following five statements:

- During my childhood, my parents allowed me to make mistakes.
- During my childhood, my parents often expressed affection and care.
- While I was growing up, I was physically punished by my parents.
- While I was growing up, my parents encouraged me to make my own decisions.

- While I was growing up, my parents emphasized the importance of obedience and compliance.

For each statement, fathers choose one of three options: *no*, *neutral*, or *yes*. These responses are coded from 1 to 3, respectively. Following the questionnaire design, we construct two indices:

- **Paternal grandparents' authoritative parenting**: the sum of items 1, 2, and 4 (range: 3–9). Higher values indicate that the paternal grandparents were perceived as more affectionate, autonomy-supportive, and tolerant of mistakes.
- **Paternal grandparents' authoritarian parenting**: the sum of items 3 and 5 (range: 2–6). Higher values indicate that the paternal grandparents were perceived as more punitive and more focused on obedience and compliance.

We use an analogous approach to construct **Maternal grandparents' authoritative parenting** and **Maternal grandparents' authoritarian parenting**, based on the mother's retrospective reports about how her parents (i.e., the child's maternal grandparents) raised her during her childhood, using the same items, response options, coding, and aggregation rules.

# D Summary statistics for all sample

Table D1: Summary Statistics of Variables from Children's Reports (All Sample)

| | Obs. (1) | Mean (2) | S.D. (3) | Min (4) | Max (5) |
|---|---|---|---|---|---|
| *Panel A. Children's characteristic* | | | | | |
| Boy (1=yes) | 3,552 | 0.53 | 0.499 | 0 | 1 |
| Age | 3,552 | 9.874 | 0.641 | 9 | 12 |
| Rural (1=yes) | 3,552 | 0.764 | 0.425 | 0 | 1 |
| Non-minority (1=yes) | 3,552 | 0.977 | 0.148 | 0 | 1 |
| # of siblings | 3,552 | 1.471 | 1.063 | 0 | 5 |
| # of illnesses during the last semester | 3,552 | 1.497 | 1.473 | 0 | 7 |
| *Panel B. Children's outcomes* | | | | | |
| Chinese test score | 3,552 | 0 | 0.998 | -5.102 | 2.374 |
| Math test score | 3,552 | 0 | 0.998 | -5.041 | 2.958 |
| Mental health | 3,552 | 8.923 | 5.272 | 0 | 30 |
| Delay of gratification | 3,552 | 0.628 | 0.483 | 0 | 1 |
| Risk preference | 3,552 | 0.295 | 0.456 | 0 | 1 |
| Altruism | 3,552 | 3.62 | 1.494 | 0 | 8 |
| Sefl-control | 3,552 | 11.788 | 3.196 | 4 | 20 |
| SDQ: | | | | | |
| Prosocial behavior | 3,552 | 4.648 | 1.101 | 2 | 6 |
| Internalizing behavior problems | 3,552 | 8.99 | 1.72 | 4 | 12 |
| Externalizing behavior problems | 3,552 | 9.425 | 1.683 | 4 | 12 |
| Big-Five: | | | | | |
| Extroversion | 3,552 | 8.802 | 2.551 | 2 | 14 |
| Agreeableness | 3,552 | 9.576 | 2.399 | 2 | 14 |
| Conscientiousness | 3,552 | 7.865 | 2.6 | 2 | 14 |
| Neuroticism | 3,552 | 8.64 | 2.69 | 2 | 14 |
| Openness | 3,552 | 8.5 | 2.491 | 2 | 14 |
| Growth mindest | 3,552 | 11.452 | 2.959 | 3 | 15 |
| Creativity | 3,552 | 26.208 | 3.924 | 14 | 36 |
| # of misbehaviors during the last week | 3,552 | 30.881 | 4.336 | 14 | 40 |
| # of nominated friends | 3,552 | 5.193 | 2.525 | 0 | 8 |
| Perceived current peer relationship quality | 3,552 | 3.485 | 1.002 | 0 | 4 |
| # of parents' verbal abuse and physical punishment (last semester) | 3,552 | 7.133 | 7.024 | 0 | 28 |
| Need to express oneself to parents | 3,552 | 2.888 | 1.069 | 1 | 4 |
| Perceived parental misunderstanding | 3,552 | 2.139 | 1.123 | 1 | 4 |
| Need for parental listening | 3,552 | 2.88 | 1.035 | 1 | 4 |
| *Panel C. Parents' parenting style* | | | | | |
| Father: | | | | | |
| Authoritative | 3,552 | 10.533 | 3.069 | 3 | 15 |
| Authoritarian | 3,552 | 6.041 | 3.066 | 3 | 15 |

*Continued on next page*

*(Continued) Summary Statistics*

| | Obs. | Mean | S.D. | Min | Max |
|---|---|---|---|---|---|
| Permissive | 3,552 | 7.99 | 2.997 | 3 | 15 |
| Neglectful | 3,552 | 4.657 | 2.201 | 2 | 10 |
| Mother: | | | | | |
| Authoritative | 3,552 | 10.936 | 3.145 | 3 | 15 |
| Authoritarian | 3,552 | 6.14 | 2.998 | 3 | 15 |
| Permissive | 3,552 | 8.049 | 3.089 | 3 | 15 |
| Neglectful | 3,552 | 4.674 | 1.777 | 2 | 10 |
| *Panel D: Child's daily time spent (in minutes)* | | | | | |
| Doing homework, reviewing, and previewing lessons | 3,552 | 43.272 | 30.214 | 0 | 120 |
| Extracurricular reading | 3,552 | 24.761 | 21.761 | 0 | 120 |
| Using Internet and playing video games | 3,552 | 21.482 | 24.849 | 0 | 120 |
| Watching TV | 3,552 | 26.995 | 27.344 | 0 | 120 |
| Doing sports/physical activities | 3,552 | 29.409 | 27.522 | 0 | 120 |
| Playing with friends/classmates | 3,552 | 37.663 | 30.555 | 0 | 120 |
| Playing alone | 3,552 | 28.394 | 28.735 | 0 | 120 |
| Studying with parental supervision | 3,552 | 28.48 | 29.758 | 0 | 120 |
| Playing with parents | 3,552 | 16.676 | 24.988 | 0 | 120 |
| Listening to stories told by parents | 3,552 | 19.291 | 24.583 | 0 | 120 |

*Notes*: *Delay of gratification* is a binary variable equal to 1 if the child chooses the delayed payment and 0 if the child chooses the immediate payment. *Risk preference* is a binary variable equal to 1 if the child chooses to flip a coin (RMB 10 if heads and RMB 0 if tails), and 0 if the child chooses to receive RMB 5 for sure. Mental health is measured using the CES-D 10 scale and is constructed by summing scores across the ten items; a higher total score indicates more severe depressive symptoms. *Altruism* is measured using a dictator game and equals the number of candies (0–8) the child chooses to give to an anonymous classmate; higher values indicate greater altruism. *Self-control* is an index constructed from four Likert-scale items on planning, time management, impulsivity, and acting without thinking. Higher values indicate higher self-control. *Prosocial behavior* is an index capturing empathy and willingness to help others; higher values indicate more prosocial behavior. *Internalizing behavior problems* is constructed as the sum of the emotional and peer-problem components. *Externalizing behavior problems* is constructed as the sum of the conduct and attention components. Higher values indicate more severe behavior problems. *Growth mindset* is a continuous index constructed by summing responses to seven 5-point Likert-scale items; higher values indicate a stronger belief that abilities (e.g., intelligence) can be improved through effort. *Creativity* is a continuous index constructed by summing responses to nine 4-point Likert-scale items; higher values indicate greater creativity. See Section C.1 for details.

Table D2: Summary Statistics of Variables from Parents' Reports (All Sample)

| | Obs. (1) | Mean (2) | S.D. (3) | Min (4) | Max (5) |
|---|---|---|---|---|---|
| *Panel A. Parents' characteristics* | | | | | |
| Father's age | 3,552 | 39.342 | 4.555 | 30 | 56 |
| Mother's age | 3,552 | 36.896 | 4.48 | 28 | 54 |
| Father' years of schooling | 3,552 | 11.18 | 3.342 | 3 | 19 |
| Mother' years of schooling | 3,552 | 11.122 | 3.378 | 3 | 19 |
| Father employed in the local county (1=yes) | 3,552 | 0.401 | 0.49 | 0 | 1 |
| Mother employed in the local county (1=yes) | 3,552 | 0.646 | 0.478 | 0 | 1 |
| *Panel B. Parents' AI usage and perceived role in parenting* | | | | | |
| Using AI for child education (1=yes) | 3,552 | 0.063 | 0.243 | 0 | 1 |
| Using AI for own study/work (1=yes) | 3,552 | 0.238 | 0.426 | 0 | 1 |
| Using AI for other purposes (1=yes) | 3,552 | 0.114 | 0.318 | 0 | 1 |
| Perceived role of AI in parenting | 3,552 | 19.292 | 5.367 | 4 | 28 |
| *Panel C. Parents' evaluation of their children* | | | | | |
| Satisfaction with the child's academic performance in the last semester | 3,552 | 7.037 | 1.922 | 0 | 10 |
| Rating of the child's academic performance in the last semester | 3,552 | 76.415 | 18.749 | 8 | 100 |
| Expectation for the child's academic achievement | 3,552 | 2.66 | 0.508 | 1 | 3 |
| *Panel D. Parents' outcomes and their parenting practices* | | | | | |
| Mental health | 3,552 | 8.36 | 4.397 | 0 | 27 |
| # of parent–child conflicts during the last semester | 3,552 | 17.741 | 14.211 | 0 | 70 |
| Parenting skills in the last semester: | | | | | |
| Rule enforcement | 3,552 | 8.124 | 1.457 | 2 | 10 |
| Emotional regulation | 3,552 | 7.522 | 1.448 | 2 | 10 |
| Communication skills | 3,552 | 7.847 | 1.678 | 2 | 10 |
| Positive parenting practices | 3,552 | 15.382 | 3.153 | 4 | 20 |
| Parenting self-efficacy | 3,552 | 21.664 | 4.311 | 6 | 30 |
| Parenting awareness | 3,552 | 29.987 | 8.086 | 8 | 40 |
| Punitive discipline measures | 3,552 | 1.01 | 1.562 | 0 | 9 |
| Command-style discipline measures | 3,552 | 1.291 | 1.721 | 0 | 9 |
| Permissive discipline measures | 3,552 | 0.381 | 0.964 | 0 | 9 |
| Communicative discipline measures | 3,552 | 6.347 | 2.757 | 0 | 9 |
| Empathy | 3,552 | 14.192 | 3.008 | 4 | 20 |
| *Panel E. Monthly Expenditures and income (in CNY)* | | | | | |
| Monthly spending on children's school supplies | 3,552 | 70.463 | 75.679 | 0 | 300 |
| Monthly spending on extracurricular resources | 3,552 | 76.566 | 79.059 | 0 | 300 |
| Monthly spending on toys | 3,552 | 53.506 | 75.674 | 0 | 300 |
| Monthly household income | 3,552 | 6,746.93 | 3,767.81 | 250 | 14,000 |
| *Panel F. Paternal and maternal grandparents' parenting styles during the parents' childhoods* | | | | | |
| Paternal grandparents: | | | | | |
| Authoritative | 3,552 | 6.396 | 1.53 | 3 | 9 |

*Continued on next page*

*(Continued) Summary Statistics*

| | Obs. | Mean | S.D. | Min | Max |
|---|---|---|---|---|---|
| Authoritarian | 3,552 | 4.417 | 1.232 | 2 | 6 |
| Maternal grandparents: | | | | | |
| Authoritative | 3,552 | 6.671 | 1.548 | 3 | 9 |
| Authoritarian | 3,552 | 4.149 | 1.256 | 2 | 6 |

*Notes*: *Mental health* is measured using the CES-D 10 scale and is constructed by summing scores across the ten items; a higher total score indicates more severe depressive symptoms. *Perceived role of AI in parenting* is an index constructed by summing four 7-point Likert-scale items on parents' attitudes toward AI tools in parenting; higher values indicate a more positive perceived role of AI in communication, understanding the child, educational advice, and parent–child relationship quality. *Parenting skills* are measured using six 5-point scenario-based items and summarized into three indices—*rule enforcement*, *emotional regulation*, and *communication skills*—each constructed by summing two items. All indices are coded so that higher values indicate better parenting skills in the corresponding dimension. *Positive parenting practices* is an index constructed by summing four 5-point Likert-scale items on supportive and emotionally responsive parenting behaviors; higher values indicate more positive parenting practices. *Parenting self-efficacy* is measured as the sum of six 5-point Likert-scale items; higher values indicate greater confidence and satisfaction in the parenting role. *Parenting awareness* is measured as the sum of eight 5-point Likert-scale items; higher values indicate greater awareness of effective parenting and the potential harms of harsh discipline and overindulgence. *Punitive discipline measure*, *command-style discipline measure*, *permissive discipline measure*, and *communicative discipline measure* are constructed from nine discipline vignettes and equal the number of situations (0–9) in which the parent reports using the corresponding practice; higher values indicate more frequent use of that practice. See Section C.2 for details.