Helping Parents Combat Middle-School Blues: Evidence from a Cluster Randomized Controlled Trial on Empathy and Parental Involvement

Qinyou Hu* Yiming Xia†

ABSTRACT:

"Middle-School Blues"— a situation in which parental mental health declines to its lowest point during their children's middle school years—is widespread yet often overlooked. Deteriorating parental mental health disrupts effective parenting and negatively impacts children's outcomes. However, there is limited research on how to improve parental mental health. We fill this gap using data from a unique cluster randomized controlled trial conducted in Chinese middle schools. In this trial, parents in the treatment classes were offered a four-month parental involvement program on empathy and positive parenting. We find that the program led to a 0.17 standard deviation increase in parental mental health for treated parents, as measured by the GHQ-12 score. The effect size indicates that the program is significantly effective in helping parents combat Middle-School Blues. We attribute the program's impact on parental mental health to improvements in parenting skills, increased time investments, better children's mental health, and the development of positive personality traits. Notably, at a cost of only $30 per class, our results indicate that this program is highly cost-effective.

*Department of Economics, Aalto University and The University of Texas at San Antonio. Email: hdfhqy@gmail.com.

†Research Institute of Economics and Management, Southwestern University of Finance and Economics. Email: xiaym@swufe.edu.cn.

KEYWORDS: Parental involvement; Mental health; Parent-child interaction; Randomized controlled trials.

JEL CLASSIFICATION: I20; J13; J24; C93

1 Introduction

According to the World Health Organization (2022), mental health is increasingly recognized as a pivotal pillar of global development. Mental illness imposes substantial costs on individuals, families, and society. Among various mental illnesses, Greenberg et al. (2021) highlights that depression alone imposes significant economic burdens, including both healthcare expenses and productivity losses, with an annual economic burden on society in the U.S. of $326.2 billion as of 2018. Notably, the costs associated with depression are primarily "indirect," with 35% classified as direct costs and 67% attributed to workplace costs.[1] The overall economic costs of parental mental health, especially maternal mental health, can be substantial and enduring. In response to these findings, there is a growing call for increased investments in addressing mental health issues (Patel et al., 2018).

Extensive literature has shown that parents' mental illnesses can cause family dysfunction and result in adverse childhood experiences, impeding children's development (Felitti et al., 1998; Nomaguchi and Milkie, 2020; Radicke et al., 2021). Nevertheless, research on parental mental health remains underexplored. The challenges of adolescence for children are well acknowledged, prompting investigations into the cost-effectiveness of interventions for children's well-being (Mihalopoulos et al., 2011; Simon et al., 2012; Mihalopoulos, 2015). However, it is noteworthy that parental mental health during this critical period has received limited recognition until recent studies highlighted the "Middle-School Blues" issues— where parental mental health reaches its lowest point when their children enter middle school—among mothers in the United States (Luthar and Ciciolla, 2015, 2016). Using nationally representative data in China, we also document a "V-shaped" pattern of parental mental health throughout children's developmental stages, highlighting the existence of "Middle-School Blues" in societies outside the United States. Despite considerable

[1]For a more detailed discussion on "direct" and "indirect" costs, see Colton and Manderscheid (2006) and McGovern et al. (2022).

attention devoted to understanding the consequences of parental mental health on their children (e.g., Le and Nguyen, 2018), practical support for parents remains limited, and the need for a scalable treatment intervention is compelling.

In this paper, we propose that a parental involvement program on empathy and positive parenting could remedy parents' "Middle-School Blues," based on evidence from a cluster randomized controlled trial (RCT) in two middle schools in Yongkang County, China. We extend the analysis of the RCT conducted by Cunha et al. (2023) to focus specifically on parents. The program was designed to tackle school bullying by developing students' empathy skills through family education, requiring parents to engage with empathy materials alongside their children. Their study provided the first comprehensive analysis of the program's effects on students' empathy and bullying reduction—the program not only successfully reduced school bullying but also improves students' outcomes, parental involvement, and parenting skills. Previous literature shows that these factors are all crucial for parental mental health.[2] Therefore, we propose that the program may also generate unintended positive effects on parental mental health, which is the focus of this paper.

We enrolled over 2,200 seventh and eighth graders and their parents from 48 classes in two middle schools, which were randomly assigned to the treatment (26 classes) and the control groups (22 classes). We then encouraged the parents and children in the treatment group to register for an online program designed to educate them on empathy and positive parenting. The program consisted of eight biweekly parent-child reading tasks on empathy, the value of uniqueness, positive parenting, and four empathy-themed movies. It encouraged parents and children to engage in empathy education both extensively and intensively. It also encouraged parents to integrate empathy into their daily parenting practices.

We collected valuable data on parental outcomes. We obtained 1,852 valid responses

[2]For example, Glatz and Buchanan (2015) suggests that parents have lower parental self-efficacy and feel more anxious about their own parenting skills due to their kids becoming harder to handle as they enter adolescence. Besides, parents also experience a higher level of education anxiety.

from parents in the follow-up survey, resulting in a moderate attrition rate of about 17%. Parents of more vulnerable students, such as those with poorer baseline mental health, were more likely to drop out, but this attrition was not correlated with the treatment modality. Regarding the program take-up rate, 41% of parents in the treatment group completed at least half of the reading or movie tasks throughout the intervention. Parents of students with better baseline outcomes were more likely to participate in the treatment, while those with children who were bullying victims or children with poorer baseline mental health were less likely to participate.

Intention-to-treat estimates suggest that the program significantly improves parental mental health, measured by the 12-item General Health Questionnaire (GHQ-12), with an increase of 0.17 standard deviations (SD). The effect is substantial enough to help parents combat the "Middle-School Blues," as indicated by a nationally representative sample showing a 0.10 SD reduction and mitigating the dip in the "V-shaped" pattern. Using random assignment as an instrument, treatment-on-the-treated estimates show that taking up the program (i.e., completing at least half of the tasks) leads to a 0.4 SD increase in GHQ-12 scores. The program performs well across all dimensions of the GHQ-12, particularly in reducing anxiety and social dysfunction. Impact heterogeneity analysis suggests that the program improves mental health for most parents in a similar manner.

With extensive measures collected on parenting skills, time investment, and children's social-emotional skills, we further explore the potential mechanisms. We attribute the program's effect on parental mental health to four mechanisms. First, the intervention improves parenting skills. Parents in treated classes have higher responsiveness and empathy. They are also more likely to adopt a democratic parenting style. Second, the program nudges parents to invest more time without displacing monetary investments. Our results also support that increased parental involvement is a positive experience for parents. Third, improvement in children's mental health enhances parental well-being. We find that students in the treatment classes report lower stress levels and are less likely to be depressed; they also

experience greater feelings of happiness following the intervention. Fourth, the intervention improves two non-cognitive abilities in students: empathy skills and positive personality traits, including self-satisfaction, self-worth, self-confidence, self-esteem, and perseverance. This finding is in line with existing research showing that parental involvement is strongly associated with reduced childhood anxiety and that improving parenting skills can help enhance children's well-being and prevent behavioral problems over the long run (Wood et al., 2003; Herring et al., 2006; Wolicki et al., 2021; Furlong et al., 2012).

Finally, we conduct a cost-effectiveness analysis for our program using the cost-benefit ratio. The program costs $30 per class, and the most conservative estimate indicates that it prevents 1.4 cases of depression per class. This translates to a cost of about $21.43 to prevent one case of depression during the four-month intervention period, which is less than the cost of a single one-on-one psychological counseling session in China, as noted by Lin (2018) (ranging from $43 to $100). Our digitally delivered parental involvement program achieves a similar effect size to internet-delivered cognitive behavioral therapy (iCBT) and peer support interventions but at a lower cost. Given that the estimated annual cost of mental health in China was $3,665 in 2013 (Xu et al., 2016), the four-month expense is higher than $1,222 in 2021. A simple back-of-the-envelope calculation estimates that our program costs $1,400, resulting in an estimated benefit of $6,734 in direct medical cost savings over four months for the 26 treatment classes. Including direct non-medical and indirect costs, the total benefit could reach as high as $44,473.

Our study contributes to several strands of literature. First, it adds to the happiness literature, particularly on the "U-shaped" happiness-age curve with a midlife dip (Blanchflower and Oswald, 2008; Cheng et al., 2017; Graham and Ruiz Pozuelo, 2017; Blanchflower, 2021), by examining parental happiness across child age and uncovering a "V-shaped" pattern in parental well-being. It also complements the developmental psychology research of Luthar and Ciciolla (2015, 2016) by demonstrating that "Middle-School Blues" issues are observed in China as well. Furthermore, we provide evidence from an RCT indicating that

parental involvement programs can enhance parental mental health.

Second, our study contributes to the literature on parental involvement, which primarily focuses on early childhood (Cunha and Heckman, 2008; Cunha et al., 2010; Bono et al., 2016; Del Boca et al., 2017; Attanasio et al., 2020; Barrera-Osorio et al., 2020). While investing in adolescence is crucial, it remains understudied (Dahl et al., 2018; Almond et al., 2018). We examine the effects of a parental involvement program on parental mental health, extending previous correlational studies on maternal depression and parenting behaviors with causal evidence (Lovejoy et al., 2000). Additionally, we are among the few to investigate parental mental health in the general population during their children's middle school years.[3] Our findings also reveal that parental time investment is positively correlated with mental health, offering empirical evidence for the debate on the role of time investment in economic models (Guryan et al., 2008).

Third, our study contributes to the literature on determinants of parental mental health. Most existing research often emphasizes "rigid" factors—such as demographics, cultural background, and life events—that are difficult to change in the short term (Lund et al., 2018). Studies on adult mental health typically link issues to poverty, family relationships, education, and childhood circumstances.[4] Our study emphasizes that parenting skills, parental time investments, and children's non-cognitive abilities—factors often overlooked in existing literature—also affect parental mental health. Recognizing that these potentially "malleable" factors can influence parental mental health has significant implications for ad-

[3]Most literature focuses on early childhood (Patterson et al., 2002; Barlow et al., 2003; Furlong et al., 2012; Barlow et al., 2012) or on specific populations, such as parents of children with developmental disorders (Tonge et al., 2006; Wyatt Kaminski et al., 2008; Aman et al., 2009; Spreckley and Boyd, 2009).

[4]Refer to Haushofer and Fehr (2014); De Quidt and Haushofer (2017); Ridley et al. (2020); Haushofer et al. (2020); Angelucci and Bennett (2021) for mental health and poverty, Thyen et al. (1999); Koball et al. (2010); Chen and van Ours (2022); Zamarro and Prados (2021) for family relationships, Janke et al. (2020); Jiang et al. (2020) for education, Almond et al. (2018) for childhood circumstances, and Böckerman et al. (2023) for parental death.

dressing the "Middle-School Blues."

Our findings have significant policy implications. The analysis indicates that the program is effective for most parents, making it applicable to societies where parental time investment is low or undervalued. The benefits to parental mental health suggest that many parental involvement programs may be more valuable and cost-effective than evaluations focused solely on children's outcomes imply. Lastly, our use of an online platform offers cost-saving benefits and improves scalability, highlighting social media as a powerful tool.

The paper is organized as follows: Section 2 introduces the "Middle-School Blues" and experiment background. Section 3 details the program and data. Section 4 explains the empirical strategy. Section 5 presents the results. Section 6 concludes.

2 Background

2.1 Parental Mental Health by Children's Developmental Stages

American mothers are prone to the "Middle-School Blues." Luthar and Ciciolla (2015, 2016) suggest that the life satisfaction of mothers reaches its nadir, and their negative emotions reach their peaks during their children's middle school period. However, related empirical evidence from other societies, particularly developing countries, is scarce.

To complement the above two studies, we first document the patterns of parental mental health status throughout child developmental stages using data from the first wave of the China Family Panel Study (CFPS).[5] Mental health is measured with the Kessler 6 Rating Scale (K-6) (Kessler et al., 2002). We sum all K-6 items to generate a total K-6 score (0-24) and use a score of five as the cutoff value to create an indicator for depression. We also control for the age to avoid confounding with the "midlife dip" as discussed in Blanchflower

[5]CFPS is a nationally representative dataset containing rich information on demographics, parent-child relationships, and parental mental health status (Xie and Hu, 2014). We use the first wave (2010) because it is sufficient to demonstrate the pattern and preserve representativeness (no attrition).

and Oswald (2008) and Blanchflower (2020), ensuring that our results specifically relate to having a middle school-aged child.

Figure 1 shows the relationship between mental health and child development stages. The y-axis displays parents' age-adjusted mental health index, and the x-axis represents children's developmental stages. As children enter middle school, parents' age-adjusted mental health index reaches its nadir, while age-adjusted depression peaks. Results from various methods confirm that the "Middle-School Blues" is not merely a result of the "midlife dip." Adjusting for the number and age of any dependent children leaves the pattern unchanged.[6] Consistent with Luthar and Ciciolla (2016), we find that, compared with fathers, mothers are more likely to experience the "Middle-School Blues" during this period.[7]

2.2 Sample Middle Schools and Sample Parents

Having demonstrated the general pattern of the 'V-shaped' relationship between parental mental health and children's developmental stages, we now proceed to discuss the characteristics of the sample in our experiment. The study sample includes seventh and eighth graders and their parents in two middle schools in a small town named Yongkang—a county belonging to Jinhua City, Zhejiang Province. Yongkang has accommodated about 0.9 million residents and achieved a GDP per capita of 103,163 RMB (about 15,000 USD) in 2020, according to the National Bureau of Statistics of China. Despite households in Yongkang

[6]Appendix Figure C1 replicates this plot, while adding the 90% confidence intervals. The same pattern is observed in Appendix Figure C2, where we replot the relationship using parents of infants (aged 0-2) as the comparison group, and control for a rich set of demographic variables and county fixed effects. Appendix Table C1 presents the regression estimates and again indicates that parents experience a statistically significant decrease (greater than 0.10 SD) in their mental health when their children enter middle school.

[7]Appendix Figure C3 suggests that fathers are generally happier than mothers across children's developmental stages. The decline in parental mental health during the middle school stage is more salient among mothers (0.13 SD) compared with fathers (0.06 SD).

being more affluent than those in other counties in China, students in the area still experience high levels of education anxiety due to the test-oriented education system, similar to most regions of China (Dello-Iacovo, 2009). Parents are also more prone to experiencing burnout due to their high expectations for their children's educational attainment and more detachment from their children. Our study sample serves as the epitome of the "Middle-School Blues" that happens among parents of "small-town swot," as reported in the magazine *The Economist (2021)*.

Another notable aspect of our sample is the low level of parental involvement. Appendix Table C3 shows that baseline parental time investments are relatively low: 40% of students reported that their parents spent 0 hours checking homework, and 42% reported 0 hours on outdoor activities in the past week.[8] This detachment may worsen mental health issues for both parents and their children. In our sample, over 20% of parents and 30% of students in the control classes can be classified as depressed.

3 The Program and Data

3.1 The Parental Involvement Program

Content

The program is a parent-directed intervention designed to enhance parenting skills and students' empathy. The intervention contains four monthly themes on empathy, perspective taking, the value of various personalities, and the role of social-emotional skills in maintaining relationships with others, as listed in Appendix Figure C4 and Table C2. The content of reading tasks mainly relies on two books written by American psychologists: *"The Power of*

[8] Appendix Table C11 shows that, on average, parents in the control group checked homework less than twice a week, watched TV together once, engaged in outdoor activities 1.4 times, and spent 3 times eating and 3.5 times discussing school life.

Empathy: A Practical Guide to Creating Intimacy, Self-understanding, and Lasting Love in Your Life citation" by A. Ciaramicoli and K. Ketcham (Ciaramicoli and Ketcham, 2000) and *"The Stress Solution: Using Empathy and Cognitive Behavioral Therapy to Reduce Anxiety and Develop Resilience"* by A. Ciaramicoli (Ciaramicoli, 2016). Each month, we deliver two reading tasks and one movie task, which require parents and students to participate together for about 30 minutes and 2 hours, respectively.

As documented in the AEA RCT Registry (Cunha et al., 2021), the initial goal of this intervention was to reduce bullying by encouraging parents to support the development of students' empathy. According to Cunha et al. (2023), who examined data following the AEA RCT Registry, the program effectively improved the empathy and positive personality traits of treated students, which subsequently led to a reduction in bullying incidents in the treated classes.

This study considers parental mental health, one of the registered variables, as the primary outcome, while the remaining registered variables—particularly those related to children—are treated as potential mechanisms for analysis. Although improving parental mental health is not the primary goal, the intervention includes extensive content on positive parenting, particularly in the first two monthly themes. For instance, week 3 introduces positive parenting skills and empathy, while week 5 discusses the effect of perspective-taking on the parent-child relationship. Additionally, we collected mental health measures for parents, which allows us to evaluate whether this intervention can improve parental mental health. In practice, these components of our intervention encouraged parents to increase meaningful parent-child interactions, fostering empathy and improving parenting skills for both parents and students. Therefore, we expect that these intermediate outcomes could also lead to an improvement in parental mental health.

Delivery

We utilized social media to deliver the intervention by incorporating the program into

a platform embedded within *WeChat*[9] and delivering the materials through this platform. Parents in randomly selected 26 classes were notified to enroll in this program. For each monthly theme, we provided two biweekly articles and one movie task. The first biweekly article focused on explaining the concept and the importance of the topic, while the second article offered real-life examples and detailed "*How-to*" procedures to educate both students and parents. For instance, the week 1 article introduced the concept of empathy and the potential benefits of being empathetic, while the week 3 article taught parents to incorporate empathy in parenting and positive parenting techniques. Regarding the movie tasks, at the beginning of each treatment month, parents received a link with instructions on how to access and watch the movie with their children. All the movies focused on themes of positive parenting or empathy.

Using social media can increase the scalability of the program for two reasons. First, it mitigates concerns about costs, as the use of social media substantially reduces the organizational expenses associated with this intervention. Second, it relies less on class teachers or school-level resources. Secondary school teachers tend to resist taking extra responsibilities, as many of the teachers suffer from some degree of burnout (Bottiani et al., 2019). In our case, during the recruitment period, teachers reported that they had no incentives to make further efforts except for forwarding biweekly messages, suggesting that family role is more important for non-cognitive skill development during the secondary school period.

3.2 Experimental Design

Figure 2 shows the timeline of the experiment. After collecting baseline student data, we randomly assigned classes to either treatment or control conditions within each stratum. Our sample includes four strata: seventh and eighth graders in the public school, "top classes," and normal-track classes (eighth graders) in the private school. Each class is a cluster in

[9] *WeChat*, similar to Facebook or WhatsApp in the United States, is the most popular social networking platform in China.

this study. Within each stratum, we randomized classes at a uniform rate of either $m/2$ if the number of classes in the stratum, m, was even or $(m+1)/2$ if m was odd. Parents and students in 26 treated classes were selected to gain access to the program.

In early March, parents in the treatment classes received a message to register on the platform and begin the first month's tasks. Over the next four months, each task message was delivered by teachers via *WeChat* groups on Friday evenings, biweekly. The teacher's role was a "messenger" in this program. In practice, records show that none of the teachers viewed any materials. Teacher non-engagement minimizes bias from teachers in the evaluation. Parents and students in the control classes received no information during the intervention. In late June 2021, parents and students in both groups received an invitation to a follow-up survey.

3.3 Data

The primary analysis in this paper uses parent survey data collected after the intervention. To collect parents' responses, we sent an invitation link to parents via class teachers' *Wechat* groups for each class. We did not collect the baseline parent survey data to avoid overwhelming tasks for parents. We supplement the analysis with student survey data. Students completed both the baseline and follow-up surveys using Computer Assisted Self-interviewing (CASI), assisted by class teachers in computer labs on campus. All participating students completed the two surveys. More than 80% of parents ($n = 1,852$) completed the follow-up survey. We provide more details regarding sample attrition in Section 3.5. To elicit unbiased answers, both parents and students were informed about the importance of confidentiality.

Primary Outcome

The primary outcome is parental mental health. We measured parental mental health using the General Health Questionnaire 12-item (GHQ-12), a widely used and validated measure for adults. GHQ-12 contains three factors: anxiety, social dysfunction, and loss

of confidence (Goldberg et al., 1997; Gao et al., 2004). However, psychology literature has shown that a unidimensional model is enough for practical research (Hankins, 2008). Therefore, we use the total score of inverted GHQ-12 (Likert scale) as the main measure, with a higher score indicating better mental health status. We also supplement it with alternative scoring methods such as GHQ scoring (0-0-1-1) and C-GHQ and use various cut-off values to indicate poor mental health (Goldberg et al., 1997; Gureje and Obikoya, 1990; Hankins, 2008).[10] For comparison purposes, we constructed standardized values relative to the control group for the three subdimensions – anxiety, social dysfunction, and loss of confidence. We also inverted the scores so that a higher score indicates better mental health status in that dimension. Additionally, we create a dummy variable for happiness based on the survey question "Do you feel happy in the past week?" It takes the value one if parents answer "Feel very happy" and zero otherwise.

Intermediate Outcomes

To understand the mechanisms of the program's impact on parental mental health, we collected comprehensive measures on parenting skills, time investment, and children's mental health and non-cognitive abilities.[11]

Parenting skills We construct a measure of parenting skills, which includes parenting style, parental responsiveness, and parents' empathy skills. We consider parenting style as one component of the parenting skills, motivated by Cobb-Clark et al. (2019), who showed that parenting style plays a critical role in children's development. To elicit parents' parenting styles, we provided detailed explanations of the four types of parenting styles – au-

[10]Preferred scoring method is Likert (0-1-2-3). The alternatives include GHQ scoring (0-0-1-1) and C-GHQ, which uses "0-0-1-1" for positive items and "0-1-1-1" for negative items.

[11]In Cunha et al. (2023), we found that the program has a precisely null effect on students' test scores, as reported again in Appendix Table C9. Hence, in the analysis of this paper, we do not consider students' test scores as one of the intermediate outcomes. However, we acknowledge that parents care about their children's academic outcomes.

thoritative, authoritarian, permissive, and neglecting – following Doepke et al. (2019), and asked the parents to select the type that best described them. We measured parental responsiveness from students' surveys; students reported the frequency with which their parents noticed their feelings and praised them. We consider parental empathy to be another important dimension of parenting skills. We used the self-reported instruments for perspective taking and empathetic concern, following Alan et al. (2021) and Kamas and Preston (2021).

Time investment The measure of time investment includes the average time spent (in hours) on parent-child activities per day during weekdays and weekends, respectively, over the past week. The activities include reading, checking homework, playing, and engaging in general education with the child. We calculate total time investment measures by summing the hours spent on all activities separately for weekdays and weekends.

Child mental health index The measures related to child mental health include stress from peers, academic outcomes, parental expectations, an indicator of depression (CES-D), and a single question about happiness.

Child non-cognitive ability index The measure of child non-cognitive abilities includes empathy skills and positive personality traits such as self-satisfaction, self-worth, self-confidence, self-esteem, and perseverance.[12] Appendix Section A.2 discusses the questions related to these measures.

Program take-up Our mobile platform tracks registration and task completion for participants. We know whether participants have viewed and downloaded each article or movie task. Since participation was voluntary, successful downloads serve as a strong indicator of task engagement. Therefore, we measure task take-up by tracking successful downloads on a biweekly basis. Additionally, we create an indicator variable for program take-up. Con-

[12]We measure students' *perspective taking*, *empathetic concern*, and *prosociality* to construct a more valid empathy measure for adolescents. This new measure of empathy aligns more closely with the modified *Interpersonal Reactivity Index* in psychology literature (Davis, 1983). We use one item each for self-satisfaction, self-worth, self-confidence, self-esteem, and perseverance.

sistent with Cunha et al. (2023), take-up is defined as completing at least half of the tasks—either four readings or two movies—during the four-month intervention. This approach allows us to investigate the effects of program participation on parents' mental health using the treatment on the treated (TOT) estimator.

3.4 Summary Statistics

Table 1 shows the summary statistics and baseline balance tests between treatment and control groups. Panel A reports statistics for parents' demographic characteristics, while Panel B reports children's characteristics. Column 1 reports the summary statistics for these characteristics in the control group, and Column 2 shows the differences in these characteristics between treatment and control groups. As shown in Panel A, parents in the study sample are, on average, 42.6 years old, with 80% of the survey respondents being mothers. Thirty percent are urban residents (*hukou*), while only 4% of them are migrants. Additionally, 93% of parents are married. Most families are relatively affluent, with just 24% of them reporting an annual family income below 100,000 RMB (approximately 15,000 USD). There are no significant differences in these demographic characteristics between the treatment and control groups.

Panel B shows that, on average, sample students are 14.5 years old, with 52% of them being male. About 47% of the students are urban residents, as indicated by their *hukou* status. Having siblings is common in this sample; only 28% of the students are only children (without siblings). Furthermore, the average height and weight of the sample are 162 cm and 50.7 kilograms, respectively. Specifically, male students have an average height and weight of 164 cm and 52.3 kilograms, while female students average 159 cm in height and 49.0 kilograms in weight. Since parents' baseline "Middle-School Blues" may be related to students' stressful feelings and poorer mental health outcomes, we have also documented the mental health status of our sample students at baseline. The average stress score is

13.2,[13] indicating that, on average, they are likely experiencing a high level of pressure. The average CES-D score is slightly below 9. Using a cutoff value of 10, 31% of the students can be classified as depressed.

3.5 Balance and Attrition

We test whether, among the non-attrited parents, the means of observable characteristics are balanced between the treatment and control groups. We examine both parents' demographics (Panel A) and their children's pre-treatment outcomes (Panel B). Column 2 in Table 1 reports the mean differences between the treatment and the control group, along with the corresponding standard errors. These differences are computed using item-wise regressions, where each variable is treated as the dependent variable, and the treatment indicator is the only regressor. Regressions control for strata fixed effects, with the standard errors clustered at the unit of randomization (class level). All 20 estimates from individual regressions are statistically insignificant, and these estimates are in line with what one would expect under a random classroom assignment and balanced attrition. Although we lack baseline measures of parental mental health, the observed balance in students' mental health between the treatment and control groups provides indirect evidence that parental mental health may also be balanced between the two groups.

The balance test serves as a preliminary check for selective attrition. To further alleviate concerns about attrition bias, we use baseline student data and conduct two main types of attrition tests (Hausman and Wise, 1979; Dumville et al., 2006): (i) a differential attrition rate test and (ii) a detailed selective attrition test, which examines the determinants of attrition. The results of these tests are presented in Appendix Table C4. The attrition variable is coded as one if the student completed the baseline survey and his/her parent did not complete the

[13] We constructed the raw stress score by combining three dimensions: (i) studies at school, (ii) peer relationships, and (iii) rank/test scores in the class. The total score ranges from 3 to 21, with higher scores indicating greater pressure levels. See more details in Appendix Section A.2.

follow-up survey and zero if the parent completed the follow-up survey.

Panel A shows that the attrition rate is 17.4%, with a difference in attrition rates between groups of 0.1%, which is not statistically significant. Panel B presents the results of the detailed analysis of attrition determinants and selective attrition tests. We find that the probability of a parent responding to the follow-up survey positively correlates with the baseline parent-child interaction levels, children's mental health, and children's positive personality traits. However, only 1 of the 19 coefficients of the interaction terms is statistically significant at 10%, suggesting that the attrition is not driven by the treatment modality.

3.6 Program Take-up

Participation in the program was purely voluntary, with parents receiving biweekly reminder messages to encourage involvement. In total, about 67.7% of eligible parents (N=680 out of 1,004) enrolled in the platform and attempted at least one task. Appendix Figure C5 shows the detailed task-completion rates for each reading and movie task. Overall, many parents were not consistent enough to complete the four-month program, and task completion rates gradually decreased. The rates increased to 60% for reading tasks and 59% for movie tasks midway through the program after teachers forwarded additional reminders. However, these rates dropped to about 22% in the final month in the final month due to preparations for final exams. Notably, only two parents in the control group accessed the platform, suggesting that the spillover problem is negligible.

Based on the previously defined program take-up, we find that 41% of eligible parents (N=412 out of 1,004) took up the program. As shown in Appendix Table C5, participation was non-random: parents of students with better mental health, less self-worth, and lower stress levels were more likely to participate. In contrast, parents of bullying victims were less likely to participate, suggesting that parents of children with better outcomes demonstrated a higher likelihood of participation.

4 Empirical Strategy

4.1 Intention to Treat

We estimate the intent-to-treat (ITT) effect by comparing outcomes across classes invited to participate in the program (treatment) and those not invited (control). We follow the empirical specification below:

$$Y_{ic} = \beta D_c + X_{ic}\delta + \epsilon_{ic}, \tag{1}$$

where Y_{ic} is a vector of outcome variables for individual i in class c at the follow-up survey, D_c is the treatment group indicator for class c, assigned at the baseline, and X_{ic} is a vector of controls that includes the strata fixed effects, along with parents' demographic characteristics, such as age, gender, residence, migration status, marital status, and income groups. These demographic characteristics are only controlled for in the robustness check. The coefficient β captures the ITT effect. For all regressions, we cluster the standard errors at the class level using the Liang-Zeger estimator. Given the small number of clusters (48), which is marginally greater than the rule of thumb, we complement it with wild cluster bootstrap (WCB) p-values using 9,999 resamplings as proposed by Cameron et al. (2008). Additionally, to leverage the experimental design, we use a permutation test with 2,000 stratified clustered resamplings (Young, 2019).

4.2 Multiple Hypothesis Testing Issue

We aim to examine the intermediate outcomes through which the program improves parental mental health. However, with 11 outcomes, we face the issue of multiple hypothesis testing (MHT). We address this issue using two approaches: dimension reduction and controlling

for familywise error rate.[14]

To reduce dimensionality, we first create an overall mechanism index variable that combines all outcomes, following Anderson (2008). Next, we categorize intermediate outcomes into three main categories: parenting skills, parental time investment, and child noncognitive ability. Using the same method, we construct inverse covariance weighting indices for these three categories. This method allows the index to capture more independent information compared to traditional factor analysis. Before aggregating all variables into the index, we inverted some outcome variables to ensure that positive effects imply an improvement in the outcome.

We also use the procedure outlined in Romano and Wolf (2005) to control familywise error rate—the probability of at least one false rejection. First, we consider outcomes at the family level and correct the p-values for those three indices. Next, we consider outcomes at the individual level and correct the p-values for individual outcomes within each family. We report the adjusted p-values for the results in Table 3.

4.3 Treatment on the Treated

The random assignment, as well as the data exported from the platform, allows us to further investigate the effects of actually participating in the program on parental mental health. We use the previously defined take-up dummy—parents who finished at least half of the family tasks are defined as successfully taking up the program. We then use the treatment assignment as the instrument for the take-up status to account for the potential self-selection issue. We estimate the treatment on the treated (TOT) effects using the following equation:

$$Y_{ic} = \lambda T_{ic} + X_{ic}\zeta + \varepsilon_{ic}, \tag{2}$$

[14] We followed the steps outlined by Bessone et al. (2021) to address the MHT problem. They also suggest a pre-analysis plan (PAP) as another strategy. Although we do not have a PAP, the experiment was preregistered with the same set of outcome variables discussed on the AEA registry.

where Y_i is a vector of outcome variables on parental mental health measures and mechanism variables, T_{ic} indicates the take-up decision of individual i in class c, and X_{ic} is a vector of controls consistent with those in equation (1). The coefficient λ captures the TOT effect. The first stage equation is:

$$T_{ic} = \gamma D_c + X_{ic}\theta + \eta_{ic}, \tag{3}$$

where random assignment dummy D_c is used as the instrument for the take-up decision.

5 Results

5.1 Impact on Parental Mental Health

Figure 3 plots the kernel densities for the preferred GHQ score across treatment and control groups. The preferred GHQ score uses a unidimensional measurement model and the Likert-scale scoring method ("1-2-3-4"). The higher the GHQ score, the better the mental health. Both distributions are slightly left-skewed. Most of the parental GHQ scores are centered around 28, indicating that the majority are moderately healthy. Compared with parents in the control group (solid line), the distribution of parents in the treatment group (dash line) shows higher skewness. The equality of the two distributions is rejected by a Kolmogorov–Smirnov test (p-value = 0.03). This figure suggests that the mental health scores of parents in the treatment group first-order stochastically dominate those in the control group.

Next, we quantify the ITT by estimating the equation (1) and report the ITT estimates in Table 2. The primary outcome is measured by the preferred GHQ score and is supplemented with a standardized GHQ score and a dummy variable for happiness.

The mean GHQ score in the control group is 28, and our ITT estimate indicates that the program resulted in a 0.79-point improvement, equivalent to 0.17 SD. Since the estimation of "Middle School Blues" in Section 2 is based on the K-6 score, we use the standardized

difference to posit our program's effect size. Relative to the 0.10-0.13 SD deterioration in parental mental health shown in Appendix Table C1, a 0.17 SD improvement in mental health induced by our program can address the "Middle School Blues" problem.

In terms of happiness, about 42% of parents in our study sample reported feeling very happy in the last week in the control group. Our program led to a 6-percentage-point increase in the likelihood of feeling very happy. All estimates are statistically significant and supported by permutation tests and the wild cluster bootstrapping method.

Parents who suffer from "Middle-School Blues" often feel more anxious, dissatisfied, and doubtful (Luthar and Ciciolla, 2015, 2016). The multiple dimensions of GHQ-12, as suggested by Gao et al. (2004), allow us to investigate the effects on its three dimensions: social dysfunction, anxiety, and loss of confidence. Those dimensions are positively coded such that a higher value indicates fewer problems in social dysfunction, anxiety, and loss of confidence. To compare the effect sizes, we standardized those scores using the mean and standard deviation of the control group. Panel B in Table 2 reports the ITT estimates for the three sub-dimensions. We find that the intervention significantly improved three dimensions of the GHQ-12. It led to a 0.12 SD increase in gaining confidence and around 0.15 SD increases in both social function and anxiety reduction. Hence, the magnitudes of the program's effects on the multiple dimensions suggest that our intervention tackles "Middle-School Blues" among Chinese parents through a comprehensive improvement in parental mental health.

Our main results are robust to attrition correction. Appendix Table C6 shows that the inverse probability weighting (Wooldridge, 2010) estimates are consistent and almost identical to the main ITT estimates. The lower and upper bounds in Lee (2009)'s bounds contain no zero, suggesting the robustness of our results to attrition.[15]

Appendix Table C7 reports the results of robustness analysis using different GHQ scor-

[15]Column 1 of Appendix Table C6 shows that the results are robust to the inclusion of control variables; the estimates remain nearly identical even when parents' demographic variables are added.

ing methods. Panel A reports the ITT estimates for GHQ-12 (Likert) and two alternative scoring methods: GHQ (0-0-1-1) and C-GHQ. All scores are standardized for comparison. The results remain fairly consistent across the different GHQ scoring methods. The GHQ-12 can be used as a screening tool for mental illness, with cut-off scores ranging from 1/2 to 6/7 of the total score, as suggested by (Goldberg et al., 1997).[16] Panel B of Table C7 reports the results using various cut-off values within this range for defining depression. Column 1 reports the mean of the variables in the control group, and Column 2 reports the ITT estimates. These ITT effects are converted into reductions in depression rates. We find that the program decreases parents' depression rates, with reductions ranging from a 3-percentage-point (marginally significant) to an 8-percentage-point. This corresponds to a 13% – 31% decrease in depression rates.

Overall, the program's effect on parental mental health is robust. The effect size, a 0.17 SD improvement, is large enough to tackle the "Middle-school Blues" problem in our study sample. This is noteworthy, given that the" Middle-School Blues" problem in the nationally representative sample features only a 0.10 SD dip in mental health (K-6 score), as shown in Figure 1. Thus, this improved mental health can help flatten the "V-shaped" curve.

Treatment on the Treated

One may also be interested in the selection into participating in the program and the associated TOT estimates. First, we find that parents of children with better baseline child outcomes were more likely to participate, as shown in Appendix Table C5. Second, we estimate the TOT effects on mental health following the equation (2) and report estimates in Column 5 of Table 2. Taking up the program, or completing at least half of the tasks, led to a 0.4 SD increase in GHQ and a 17 percentage point increase in the likelihood of feeling very happy in the past week. The effect size of TOT on standardized scores is similar to that

[16] Goldberg et al. (1997) suggest that the best cut-off score should vary from 1/2 to 6/7. Currently, no consensus has been reached among existing literature, though 2/3 is the most commonly used threshold value.

of other low-cost parental mental health support interventions, though most interventions targeting parents aim to tackle prenatal depression. For instance, the result of meta-analysis in Huang et al. (2020) suggests that peer support interventions lead to a reduction of the mean standardized depressive score by 0.37 among mothers.

It is also important to discuss whether the TOT effects are greater than the ATE effects. Current take-up patterns suggest that TOT effects are driven by parents with better baseline child outcomes, which are likely correlated with better parenting skills. Thus, there is a positive selection into the program in the treatment group, as it is a voluntary education program. ATE estimates reflect the average impact of the program, assuming that everyone participates in it. Speculatively, the comparison between TOT and ATE is theoretically unclear. On the one hand, TOT is likely to be larger than ATE as those who chose to participate are motivated, and learning outcomes depend on initial ability level (Cunha and Heckman, 2007). On the other hand, those never-takers show poorer outcomes across all dimensions, holding significant potential for improvement. In this case, ATE could be higher than TOT if the remedy effects are strong enough. However, empirically, for our study, the difference between TOT and ATE is likely small and insignificant, as we have not detected any significant heterogeneity, as discussed below.

Heterogeneous Treatment Effects

Our heterogeneity analysis focuses on two aspects: (1) parents' demographics and socioeconomic status, and (2) children's demographics and baseline mental health related outcomes.[17] Appendix Figure C7 illustrates the results for the primary outcome variables—parents' GHQ score. We observe only a few insignificant impact heterogeneity across different dimensions, with all impacts being positive for parents of diverse characteristics. Appendix Table C8 indicates that 9 out of 20 groups are still significant at the 10% level, even

[17]Specifically, family socioeconomic status is measured by household income and parents' demographic characteristics, while children's characteristics consist of age, gender, baseline stress score, and CES-D score.

after we account for multiple hypothesis testing using the procedure in Romano and Wolf (2005), suggesting that our parental involvement program suits the majority of the parents and almost all can benefit by improving their mental health.

5.2 Mechanisms

Why is middle school the toughest period for parents? Literature identifies three reasons for the "Middle-School Blues" issue. First, parents often lack the skills to navigate the challenges of adolescents in transition (Montemayor, 1983; Baumrind, 1991), which diminishes their parental self-efficacy (Glatz and Buchanan, 2015). This period is challenging for children as they transition into adolescence and encounter unfamiliar school environments. It is also a period with a heightened risk of engaging in risky behaviors (Steinberg and Silk, 2002; Luthar and Ciciolla, 2016). The growing complexity of adolescence demands greater parenting skills and involvement.

Second, parents and children gradually become distant during this period. Adolescents can change rapidly, while parents tend to rely on routine parenting. Adolescents question their parents' authority and are more likely to turn to peers for their primary sources of intimacy and emotional support (Nomaguchi, 2012). Additionally, the middle school system reduces parental involvement, as many parents feel less capable of assisting with their children's schoolwork during this time (Hill and Tyson, 2009).

Third, parents are also concerned about their children's emotional health and abilities. Literature has shown a contagion of stress from children to parents (Luthar and Ciciolla, 2016). During the transition, children must navigate new and relatively impersonal school environments (Eccles et al., 1993). As a result, children may feel more stressed during this period, and this stress can "spill up" to their parents.

Together, the literature serves as the theoretical foundation for our analysis of mechanisms, as illustrated in Appendix Figure C8. With the rich data collected, we empirically test program impacts on the three mechanisms: parenting skills, time investment, and children's

abilities.

Our sample for the mechanism study consists of the matched parent-child data (N=1,852). Table 3 reports the ITT results. First, we examine the program's impact on an overall mechanism index that aggregates all intermediate outcomes following Anderson (2008). Panel A shows that parents in the treatment group scored 0.31 SD higher on the overall mechanism index compared to those in the control group, suggesting that the program has an overall positive effect on the selected intermediate outcomes. We further examine each index and outcome in greater detail. Parents in the treatment group scored about 0.20 SD higher in parenting skills and 0.15 SD higher in time investments. Children in treated classes also scored 0.23 SD higher in mental health and 0.17 SD higher in non-cognitive ability indices.

We further examine the program's impact on each mechanism variable. Panel B reports the results for parental outcomes, including parenting skills and time investments. Regarding parenting skills, treated parents scored 0.10 SD higher on empathy, indicating improved parental skills. Additionally, our program increased the likelihood of adopting a democratic parenting style by 3.9 percentage points. According to students' reports, parents in the treated group were more likely to understand their children's feelings and to encourage hard work. In terms of time investments, our program led to an average increase of 0.51 hours on weekdays and 0.41 hours on weekends.[18]

Panel C reports the results for children's outcomes, including mental health and non-cognitive skills. Our program significantly improved students' mental health, decreasing the likelihood of experiencing depression by 4.8 percentage points. It also significantly reduced students' stress levels by 0.20 SD. Additionally, students showed significant improvements in their non-cognitive skills. Treated students gained 0.12 SD in empathy. They also experienced a 0.17 SD increase in positive personality measures, including self-satisfaction,

[18]The control group's average time investment is 3.73 hours on weekdays and 5.41 hours on weekends. Additional results on specific parent-child activities are provided in Appendix Section B.2, and further evidence on monetary investments are discussed in Appendix Section B.3.

self-worth, self-confidence, self-esteem, and perseverance. We also test the program's effect on students' test scores. However, the results indicate that the program had precisely null effects on students' test scores, as shown in Appendix Table C9.

The results are robust to alternative tests. Columns 3 – 5 in Table 3 report p-values from the alternative methods discussed in Section 4.2 to address several concerns and test the robustness of our findings: randomization design (Column 3), small clusters (Column 4), and multiple-hypothesis testing (Column 5). The impacts on all intermediate outcomes remain robust across these three methods. Specifically, to alleviate concerns about multiple hypothesis testing, we report adjusted p-values following Romano and Wolf (2005). These adjusted p-values are all below 0.10, as individual outcomes are highly correlated within each family.

Finally, Column 6 presents the TOT estimates for all mechanism variables, estimated using equation (2). Compared to ITT estimates, the effect sizes of TOT estimates are more than doubled. Treated parents exceeded the control group by 0.75 SD on the overall index. Specifically, as shown in Panel A, the TOT estimates indicate that taking up the program significantly improved parenting skills by 0.47 SD, time investment by 0.37 SD, child mental health by 0.55 SD, and non-cognitive ability by 0.41 SD. Panels B and C report the TOT estimates for each item used to construct these indices.

5.3 Cost-effectiveness Analysis

In this section, we present a simple cost-effectiveness analysis of our program and provide a cost-benefit ratio (cost per benefit) to the intervention. Table 4 breaks down the results into three panels: benefits (Panel A), organizational costs (Panel B), and opportunity costs (Panel C).

As shown in Panel A, the ITT estimates from Table 2 suggest that our program resulted in an average short-run benefit of 0.17 SD in the parental mental health index, as measured by the standardized GHQ score. Using the most conservative estimates for depression indi-

cators from Appendix Table C7, our program is estimated to prevent 1.4 cases of depression per class over the four-month duration of the program. TOT estimates suggest a 0.37 SD improvement in parental mental health for those who took up the program, preventing 3.1 cases of depression per class.[19]

As shown in Panel B, the total organizational cost of our program is $1,400 for all 48 classes. This amount includes (i) the cost of hiring a temporary research assistant to monitor the implementation process and (ii) the cost of gifts given to the 48 class teachers who helped forward reminder messages to parents. The total organizational cost ($1,400) and the cost per class ($30) are the key figures for evaluating the feasibility of scaling up the program. The cost-benefit ratios suggest that our program is cost-effective, requiring only about $21.43 to prevent one case of depression per class and approximately $18 to level up parents' mental health by 0.10 SD. The cost per parent is as low as $0.64. When accounting for the opportunity costs of voluntary parental time investments, the cost per treated parent is around $25.53, as shown in Panel C.

We also position our cost-effectiveness results within the existing literature. Panel A in Appendix Table C12 summarizes several relatively low-cost yet impactful parental mental health interventions. One caveat is that most of these studies focus on mothers' mental health and target the prenatal period. The effect size of our ITT estimate is comparable to that of internet-delivered cognitive behavioral therapy (iCBT) as reported by Richards et al. (2020), which costs £95 per patient. Delgadillo et al. (2022) suggests that incorporating a pre-screening stage and treating patients based on their mental health status can improve the

[19] Appendix Table C7 (Column (2) in Panel B) reports estimates for the reduction in depression rates using various alternative cut-off values, with the effects ranging from -0.028 to -0.077. We use the most conservative ITT estimate of a 0.028 percentage point reduction to calculate the program's impact on preventing depression incidents. Given the average class size of 47, we estimate the number of depressions averted per class as $0.028 \times 47 \approx 1.4$. Similarly, using TOT estimates, the number of depression cases averted per class is $0.066 \times 47 \approx 3.1$.

effectiveness to about 0.5 SD at a cost of £248 per patient. Another low-cost intervention, conducted by Fuhr et al. (2019), utilizes peer support intervention in India and incurs costs of $133 per patient. Thus, compared to iCBT and peer support, our digitally delivered parental involvement program achieves a similar effect size at a lower cost.

In Panel B, we present the estimated costs associated with treating mental health problems as found in the literature, which can also be interpreted as the estimated benefit. To convert the benefits of preventing depression into a dollar figure, we refer to Ding et al. (2022), which provides the closest context to our study: treating one depressed patient would save an annual direct medical cost of $555.5. The potential savings can reach as high as $3,665.4 when accounting for direct non-medical and indirect costs, as reported in Xu et al. (2016).

Related to our intervention, it costs only $21.43 to avert one case of depression during the intervention, assuming it is effective for only four months. This represents approximately one-third of the annual mental health costs, specifically $185 in medical costs and $1,222 in total costs. Considering that our intervention incurred a total of $1,400 in organizational costs while averting, on average, 1.4 cases of depression across 26 classes, this results in an estimated benefit of $6,734 direct medical cost savings over four months, which can even rise to $44,473 in total cost.[20]

We should also note that our designed platform can be easily adapted to other contexts and used repeatedly, thereby reducing the fixed costs associated with scaling up. Further-

[20]In this calculation, we assume that the short-term reduction in depression lasts about four months, representing one-third of the annual costs. Thus, the medical costs averted are calculated as: $\frac{1}{3} \times \$555,5 = \$185$; similarly the total cost is: $\frac{1}{3} \times \$3,665.4 = \$1,222$. The total benefits of our program are multiplied by 1.4 (the effect per class) and 26 treated classes. Therefore, the total savings on medical costs is: $\$185 \times 1.4 \times 26 = \$6,734$; and total savings amount to: $\$1,222 \times 1.4 \times 26 = \$44,473$. Given that some relevant education program effects faded rapidly after the implementation (Puma et al., 2010, 2012), it is reasonable and conservative to use the lower bound of the benefit (i.e., using four months and assuming that the effect diminishes to zero after the program).

more, the intervention required only teachers to forward our messages to remind parents on a monthly basis, with no monetary incentives provided. Consequently, we believe it can be effectively scaled up even in low-income contexts.

6 Conclusion

Parents, particularly those with teenagers, face the risk of experiencing "Middle-School Blues" and may struggle to connect with their adolescent children. We utilize nationally representative data in China to confirm the existence of this problem. More importantly, we draw on findings from a parental involvement program to propose a potential policy implication. Our study shows that participating in a program focused on empathy education and positive parenting can significantly improve parents' mental health during this challenging period. In exploring the mechanisms, we find that the primary reason for this improvement may stem from the improved parenting skills parents acquired through the program. Additionally, the non-cognitive abilities of their children may contribute to improving parents' well-being. Notably, parents are likely to gain utility directly from the time spent with children, viewing time investment as leisure rather than a burden.

Our study makes a step forward in understanding parental well-being, which has significant implications for child well-being, fertility outcomes, and society as a whole (Nomaguchi and Milkie, 2020). When interpreting the results, it is important to consider that our study primarily focuses on the short-term effects on parental mental health. While our findings demonstrate cost-effectiveness even with short-run benefit estimates, understanding the long-term effects is equally, if not more, crucial. Further research into the long-term impact of parental involvement programs would help assess more accurate total benefits.

Our program and findings are potentially generalizable to societies where parental time investment is low and parent-child tension is high. We discuss the scalability of our program in detail, following the four transparent SANS conditions (List, 2020). First, we se-

lect seventh and eighth graders from two schools (one public and one private) in a county. Our sample reflects parents in that county whose children face high academic pressure and whose parental time investment is low. Second, the attrition rate of the parent survey is about 17%. There is no evidence of differential attrition patterns across treatment arms. Besides, the characteristics of the study sample (non-attrited) are also balanced across treatment arms. Third, our parental involvement treatment is quite natural to parents, as many are accustomed to receiving homework tasks from the teachers via the social media app. The experimental sessions take place in a family setting, where all decisions are made spontaneously rather than imposed. Lastly, the intervention is low-cost, making it feasible to scale up and cost-effective. That said, it is essential to comprehensively examine the local context regarding parental involvement and mental health before applying these findings to other locations, particularly in more developed regions or countries that share few cultural similarities.

Acknowledgment

We would like to thank Julian Reif and two anonymous referees for helpful comments and suggestions. We are also grateful for discussions with Lindsey Bullinger, Pedro Carneiro, Flávio Cunha, Emily Lawler, Elaine Liu, Lena Song, Fengyu Wu, Anwen Zhang, and participants at the University of Houston Brown Bag Seminar, the 2022 Southern Economic Association (SEA) Annual Meeting, and the 2023 Atlanta Workshop on Public Policy and Child Well-being.

References

Alan, S., C. Baysan, M. Gumren, and E. Kubilay (2021). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *The Quarterly Journal of Economics 136*(4), 2147–2194.

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature 56*(4), 1360–1446.

Aman, M. G., C. J. Mcdougle, L. Scahill, B. Handen, L. E. Arnold, C. Johnson, K. A. Stigler, K. Bearss, E. Butter, N. B. Swiezy, et al. (2009). Medication and parent training in children with pervasive developmental disorders and serious behavior problems: results from a randomized clinical trial. *Journal of the American Academy of Child & Adolescent Psychiatry 48*(12), 1143–1154.

Anderson, M. L. (2008). 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 103*(484), 1481–1495.

Angelucci, M. and D. Bennett (2021). The economic impact of depression treatment in india. Technical report, IZA Discussion Paper No. 14393.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review 110*(1), 48–85.

Barlow, J., E. Coren, and S. Stewart-Brown (2003). Parent-training programmes for improving maternal psychosocial health. *Cochrane Database of Systematic Reviews 4*, CD002020.

Barlow, J., N. Smailagic, N. Huband, V. Roloff, and C. Bennett (2012). Group-based parent

training programmes for improving parental psychosocial health. *Campbell Systematic Reviews 8*(1), 1–197.

Barrera-Osorio, F., P. Gertler, N. Nakajima, and H. Patrinos (2020). Promoting parental involvement in schools: Evidence from two randomized experiments. Technical report, National Bureau of Economic Research.

Baumrind, D. (1991). Effective parenting during the early adolescent transition. *Family Transitions 2*(1), 1.

Bessone, P., G. Rao, F. Schilbach, H. Schofield, and M. Toma (2021). The economic consequences of increasing sleep among the urban poor. *The Quarterly Journal of Economics 136*(3), 1887–1941.

Blanchflower, D. G. (2020). Unhappiness and age. *Journal of Economic Behavior & Organization 176*, 461–488.

Blanchflower, D. G. (2021). Is happiness u-shaped everywhere? age and subjective well-being in 145 countries. *Journal of Population Economics 34*(2), 575–624.

Blanchflower, D. G. and A. J. Oswald (2008). Is well-being u-shaped over the life cycle? *Social Science & Medicine 66*(8), 1733–1749.

Böckerman, P., M. Haapanen, and C. Jepsen (2023). Dark passage: Mental health consequences of parental death. *American Journal of Health Economics 9*(4), 584–604.

Bono, E. D., M. Francesconi, Y. Kelly, and A. Sacker (2016). Early maternal time investment and early child outcomes. *The Economic Journal 126*(596), F96–F135.

Bottiani, J. H., C. A. Duran, E. T. Pas, and C. P. Bradshaw (2019). Teacher stress and burnout in urban middle schools: Associations with job demands, resources, and effective classroom practices. *Journal of School Psychology 77*, 36–51.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Chen, S. and J. C. van Ours (2022). Mental health effects of same-sex marriage legalization. *Health Economics 31*(1), 42–56.

Cheng, T. C., N. Powdthavee, and A. J. Oswald (2017). Longitudinal evidence for a midlife nadir in human well-being: Results from four data sets. *The Economic Journal 127*(599), 126–142.

Ciaramicoli, A. P. (2016). *The Stress Solution: Using Empathy and Cognitive Behavioral Therapy to Reduce Anxiety and Develop Resilience*. New World Library.

Ciaramicoli, A. P. and K. Ketcham (2000). *The Power of Empathy: A Practical Guide to Creating Intimacy, Self-understanding, and Lasting Love in Your Life*. Dutton.

Cobb-Clark, D. A., N. Salamanca, and A. Zhu (2019). Parenting style as an investment in human development. *Journal of Population Economics 32*(4), 1315–1352.

Colton, C. W. and R. W. Manderscheid (2006). Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. *Preventing Chronic Disease 3*(2), A42.

Cunha, F. et al. (2021, January 22). Cultivating youth's empathy through parental involvement: Experimental evidence from china. AEA RCT Registry. https://doi.org/10.1257/rct.7079-1.0.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Cunha, F., Q. Hu, Y. Xia, and N. Zhao (2023). Reducing bullying: Evidence from a parental involvement program on empathy education. Working Paper 30827, National Bureau of Economic Research.

Dahl, R. E., N. B. Allen, L. Wilbrecht, and A. B. Suleiman (2018). Importance of investing in adolescence from a developmental science perspective. *Nature 554*(7693), 441–450.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology 44*(1), 113.

De Quidt, J. and J. Haushofer (2017). Depression through the lens of economics: A research agenda. In *The Economics of Poverty Traps*, pp. 127–152. University of Chicago Press.

Del Boca, D., C. Monfardini, and C. Nicoletti (2017). Parental and child time investments and the cognitive development of adolescents. *Journal of Labor Economics 35*(2), 565–608.

Delgadillo, J., S. Ali, K. Fleck, C. Agnew, A. Southgate, L. Parkhouse, Z. D. Cohen, R. J. DeRubeis, and M. Barkham (2022). Stratified care vs stepped care for depression: A cluster randomized clinical trial. *JAMA Psychiatry 79*(2), 101–108.

Dello-Iacovo, B. (2009). Curriculum reform and 'quality education' in china: An overview. *International Journal of Educational Development 29*(3), 241–249.

Ding, R., D. Zhu, Y. Wang, M. Yong, X. Shi, and P. He (2022). Medical service utilisation and direct medical cost of depression: a cross-sectional analysis of urban medical claims data from china. *BMJ Open 12*(9), e056422.

Doepke, M., G. Sorrenti, and F. Zilibotti (2019). The economics of parenting. *Annual Review of Economics 11*, 55–84.

Dumville, J. C., D. J. Torgerson, and C. E. Hewitt (2006). Reporting attrition in randomised controlled trials. *BMJ 332*(7547), 969–971.

Eccles, J. S., C. Midgley, A. Wigeld, C. Buchanan, D. Reuman, C. Flanagan, and D. Mac Iver (1993). Development during adolescence. the impact of stage-environment on young adolescents experiences in schools and in families. *American Psychologist 48*(2), 90–101.

Felitti, V. J., R. F. Anda, D. Nordenberg, D. F. Williamson, A. M. Spitz, V. Edwards, and J. S. Marks (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The adverse childhood experiences (ace) study. *American Journal of Preventive Medicine 14*(4), 245–258.

Fuhr, D. C., B. Weobong, A. Lazarus, F. Vanobberghen, H. A. Weiss, D. R. Singla, H. Tabana, E. Afonso, A. De Sa, E. D'Souza, et al. (2019). Delivering the thinking healthy programme for perinatal depression through peers: an individually randomised controlled trial in india. *The Lancet Psychiatry 6*(2), 115–127.

Furlong, M., S. McGilloway, T. Bywater, J. Hutchings, S. M. Smith, and M. Donnelly (2012). Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Campbell Systematic Reviews 8*(1), 1–239.

Gao, F., N. Luo, J. Thumboo, C. Fones, S.-C. Li, and Y.-B. Cheung (2004). Does the 12-item general health questionnaire contain multiple factors and do we need them? *Health and Quality of Life Outcomes 2*(1), 1–7.

Glatz, T. and C. M. Buchanan (2015). Change and predictors of change in parental self-efficacy from early to middle adolescence. *Developmental Psychology 51*(10), 1367.

Goldberg, D. P., R. Gater, N. Sartorius, T. B. Ustun, M. Piccinelli, O. Gureje, and C. Rutter

(1997). The validity of two versions of the ghq in the who study of mental illness in general health care. *Psychological Medicine 27*(1), 191–197.

Graham, C. and J. Ruiz Pozuelo (2017). Happiness, stress, and age: How the u curve varies across people and places. *Journal of Population Economics 30*(1), 225–264.

Greenberg, P. E., A.-A. Fournier, T. Sisitsky, M. Simes, R. Berman, S. H. Koenigsberg, and R. C. Kessler (2021). The economic burden of adults with major depressive disorder in the united states (2010 and 2018). *Pharmacoeconomics 39*(6), 653–665.

Gureje, O. and B. Obikoya (1990). The ghq-12 as a screening tool in a primary care setting. *Social Psychiatry and Psychiatric Epidemiology 25*(5), 276–280.

Guryan, J., E. Hurst, and M. Kearney (2008). Parental education and parental time with children. *Journal of Economic Perspectives 22*(3), 23–46.

Hankins, M. (2008). The reliability of the twelve-item general health questionnaire (ghq-12) under realistic assumptions. *BMC Public Health 8*(1), 1–7.

Haushofer, J. and E. Fehr (2014). On the psychology of poverty. *Science 344*(6186), 862–867.

Haushofer, J., R. Mudida, and J. P. Shapiro (2020). The comparative impact of cash transfers and a psychotherapy program on psychological and economic well-being. Technical report, National Bureau of Economic Research.

Hausman, J. A. and D. A. Wise (1979). Attrition bias in experimental and panel data: the gary income maintenance experiment. *Econometrica 47*(2), 455–473.

Herring, S., K. Gray, J. Taffe, B. Tonge, D. Sweeney, and S. Einfeld (2006). Behaviour and emotional problems in toddlers with pervasive developmental disorders and developmental delay: Associations with parental mental health and family functioning. *Journal of Intellectual Disability Research 50*(12), 874–882.

Hill, N. E. and D. F. Tyson (2009). Parental involvement in middle school: a meta-analytic assessment of the strategies that promote achievement. *Developmental Psychology 45*(3), 740.

Huang, R., C. Yan, Y. Tian, B. Lei, D. Yang, D. Liu, and J. Lei (2020). Effectiveness of peer support intervention on perinatal depression: A systematic review and meta-analysis. *Journal of Affective Disorders 276*, 788–796.

Janke, K., D. W. Johnston, C. Propper, and M. A. Shields (2020). The causal effect of education on chronic health conditions in the uk. *Journal of Health Economics 70*, 102252.

Jiang, W., Y. Lu, and H. Xie (2020). Education and mental health: Evidence and mechanisms. *Journal of Economic Behavior & Organization 180*, 407–437.

Kamas, L. and A. Preston (2021). Empathy, gender, and prosocial behavior. *Journal of Behavioral and Experimental Economics 92*, 101654.

Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. Normand, E. E. Walters, and A. M. Zaslavsky (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine 32*(6), 959–976.

Koball, H. L., E. Moiduddin, J. Henderson, B. Goesling, and M. Besculides (2010). What do we know about the link between marriage and health? *Journal of Family Issues 31*(8), 1019–1040.

Le, H. T. and H. T. Nguyen (2018). The impact of maternal mental health shocks on child health: Estimates from fixed-effects instrumental variables models for two cohorts of australian children. *American Journal of Health Economics 4*(2), 185–225.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Lin, I.-T. S. (2018). The mental healthcare industry in china. Accessed: 2024-10-01. https://www.china-briefing.com/news/mental-healthcare-industry-china.

List, J. A. (2020). Non est disputandum de generalizability? a glimpse into the external validity trial. Working Paper 27535, National Bureau of Economic Research.

Lovejoy, M. C., P. A. Graczyk, E. O'Hare, and G. Neuman (2000). Maternal depression and parenting behavior: A meta-analytic review. *Clinical Psychology Review 20*(5), 561–592.

Lund, C., C. Brooke-Sumner, F. Baingana, E. C. Baron, E. Breuer, P. Chandra, J. Haushofer, H. Herrman, M. Jordans, C. Kieling, et al. (2018). Social determinants of mental disorders and the sustainable development goals: a systematic review of reviews. *The Lancet Psychiatry 5*(4), 357–369.

Luthar, S. S. and L. Ciciolla (2015). Who mothers mommy? factors that contribute to mothers' well-being. *Developmental Psychology 51*(12), 1812.

Luthar, S. S. and L. Ciciolla (2016). What it feels like to be a mother: Variations by children's developmental stages. *Developmental Psychology 52*(1), 143.

McGovern, M. E., S. Rokicki, and N. E. Reichman (2022). Maternal depression and economic well-being: A quasi-experimental approach. *Social Science & Medicine 305*, 115017.

Mihalopoulos, C. (2015). Assessing the economic costs of parental mental illness. In A. E. Reupert (Ed.), *Parental psychiatric disorder: Distressed parents and their families* (3rd ed.)., pp. 42–50. Cambridge: Cambridge University Press.

Mihalopoulos, C., T. Vos, J. Pirkis, and R. Carter (2011). The economic analysis of prevention in mental health programs. *Annual Review of Clinical Psychology 7*(1), 169–201. PMID: 21443447.

Montemayor, R. (1983). Parents and adolescents in conflict: All families some of the time and some families most of the time. *The Journal of Early Adolescence 3*(1-2), 83–103.

Nomaguchi, K. and M. A. Milkie (2020). Parenthood and well-being: A decade in review. *Journal of Marriage and Family 82*(1), 198–223.

Nomaguchi, K. M. (2012). Parenthood and psychological well-being: Clarifying the role of child age and parent–child relationship quality. *Social Science Research 41*(2), 489–498.

Patel, V., S. Saxena, C. Lund, G. Thornicroft, F. Baingana, P. Bolton, D. Chisholm, P. Y. Collins, J. L. Cooper, J. Eaton, et al. (2018). The lancet commission on global mental health and sustainable development. *The Lancet 392*(10157), 1553–1598.

Patterson, J., J. Barlow, C. Mockford, I. Klimes, C. Pyper, and S. Stewart-Brown (2002). Improving mental health through parenting programmes: block randomised controlled trial. *Archives of Disease in Childhood 87*(6), 472–477.

Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn, and J. Downer (2012). Third grade follow-up to the head start impact study: Final report (OPRE report 2012-45). U.S. Department of Health and Human Services, Administration for Children & Families.

Puma, M., S. Bell, R. Cook, C. Heid, G. Shapiro, P. Broene, F. Jenkins, P. Fletcher, L. Quinn, J. Friedman, et al. (2010). Head start impact study. Final report. U.S. Department of Health and Human Services, Administration for Children & Families.

Radicke, A., C. Barkmann, B. Adema, A. Daubmann, K. Wegscheider, and S. Wiegand-Grefe (2021). Children of parents with a mental illness: Predictors of health-related quality of life and determinants of child–parent agreement. *International Journal of Environmental Research and Public Health 18*(2), 379.

Richards, D., A. Enrique, N. Eilert, M. Franklin, J. Palacios, D. Duffy, C. Earley, J. Chapman, G. Jell, S. Sollesse, et al. (2020). A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. *NPJ Digital Medicine 3*(1), 85.

Ridley, M., G. Rao, F. Schilbach, and V. Patel (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science 370*(6522), eaay0214.

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

Simon, E., C. Dirksen, S. Bögels, and D. Bodden (2012). Cost-effectiveness of child-focused and parent-focused interventions in a child anxiety prevention program. *Journal of Anxiety Disorders 26*(2), 287–296.

Spreckley, M. and R. Boyd (2009). Efficacy of applied behavioral intervention in preschool children with autism for improving cognitive, language, and adaptive behavior: a systematic review and meta-analysis. *The Journal of Pediatrics 154*(3), 338–344.

Steinberg, L. and J. S. Silk (2002). Parenting adolescents. In M. H. Bornstein (Ed.), *Handbook of parenting: Children and parenting, Vol*, Volume 1, pp. 103–133. Lawrence Erlbaum Associates Publishers, xxxviii: Mahwah, NJ, US.

The Economist (2021). Education in china is becoming increasingly unfair to the poor. Accessed: 2024-3-08. https://www.economist.com/china/2021/05/27/education-in-china-is-becoming-increasingly-unfair-to-the-poor.

Thyen, U., K. Kuhlthau, and J. M. Perrin (1999). Employment, child care, and mental health of mothers caring for children assisted by technology. *Pediatrics 103*(6), 1235–1242.

Tonge, B., A. Brereton, M. Kiomall, A. Mackinnon, N. King, and N. Rinehart (2006). Effects on parental mental health of an education and skills training program for parents

of young children with autism: A randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry 45*(5), 561–569.

Wolicki, S. B., R. H. Bitsko, R. A. Cree, M. L. Danielson, J. Y. Ko, L. Warner, and L. R. Robinson (2021). Mental health of parents and primary caregivers by sex and associated child health indicators. *Adversity and Resilience Science 2*(2), 125–139.

Wood, J. J., B. D. McLeod, M. Sigman, W.-C. Hwang, and B. C. Chu (2003). Parenting and childhood anxiety: Theory, empirical findings, and future directions. *Journal of Child Psychology and Psychiatry 44*(1), 134–151.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

World Health Organization (2022). World mental health report: transforming mental health for all. Technical report, World Health Organization.

Wyatt Kaminski, J., L. A. Valle, J. H. Filene, and C. L. Boyle (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology 36*(4), 567–589.

Xie, Y. and J. Hu (2014). An introduction to the china family panel studies (cfps). *Chinese Sociological Review 47*(1), 3–29.

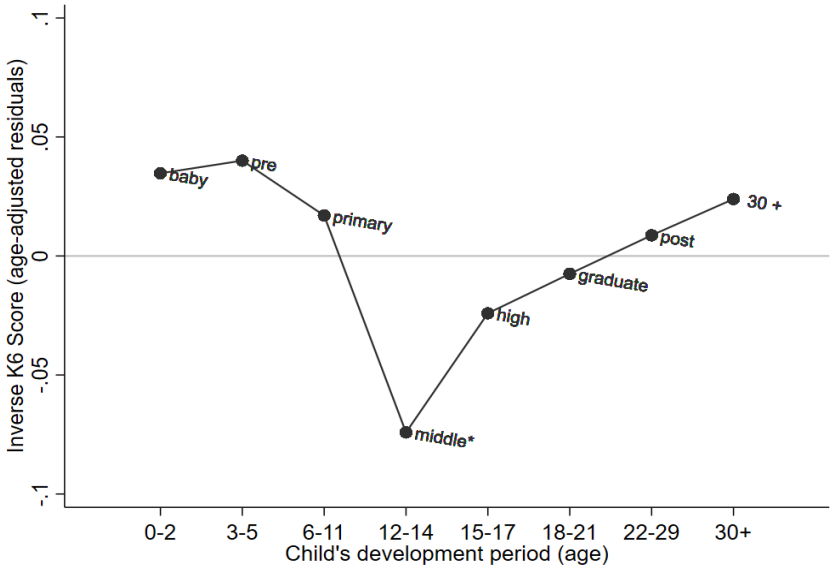
Xu, J., J. Wang, A. Wimo, and C. Qiu (2016). The economic burden of mental disorders in china, 2005–2013: implications for health policy. *BMC Psychiatry 16*, 1–9.

Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics 134*(2), 557–598.

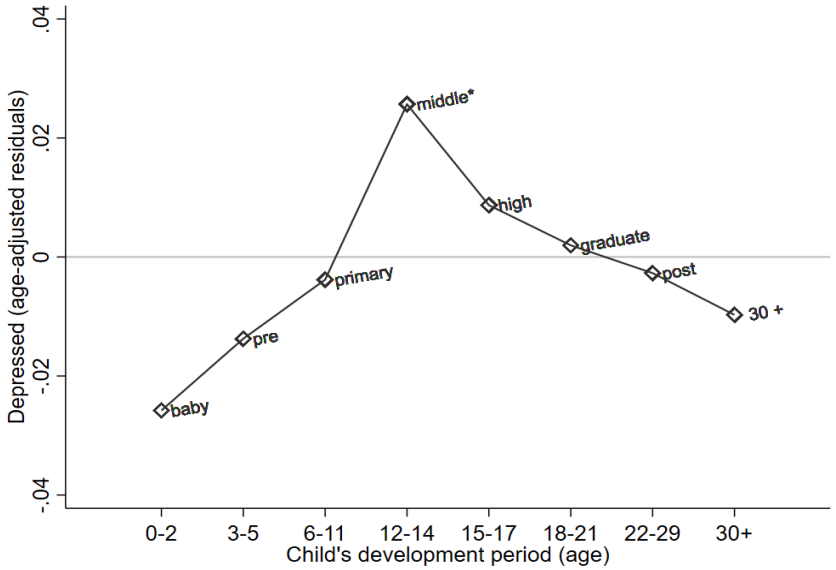
Zamarro, G. and M. J. Prados (2021). Gender differences in couples' division of childcare, work and mental health during covid-19. *Review of Economics of the Household 19*(1), 11–40.

Figure 1: Parental Mental Health Across Children's Developmental Periods, measured by K-6 Score

Panel A. K-6 Score

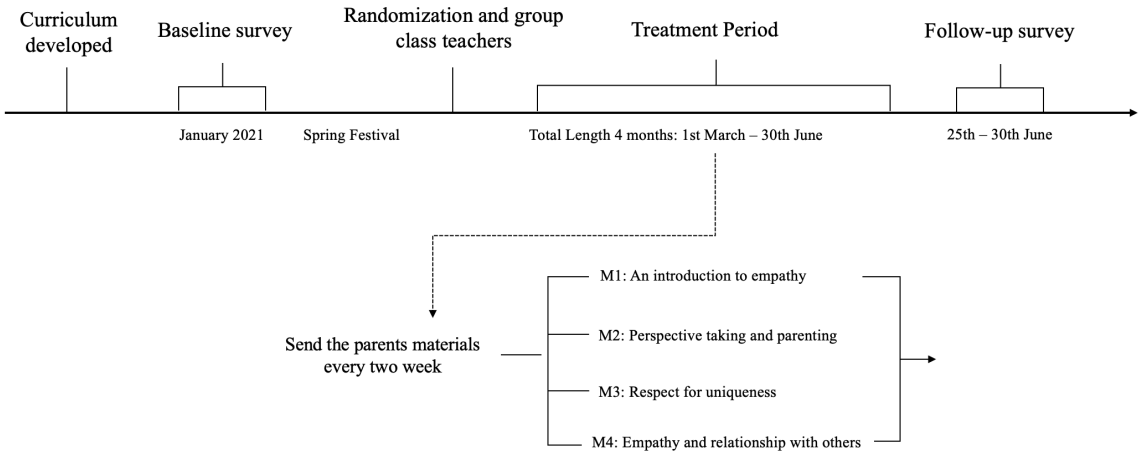


Panel B. Depression



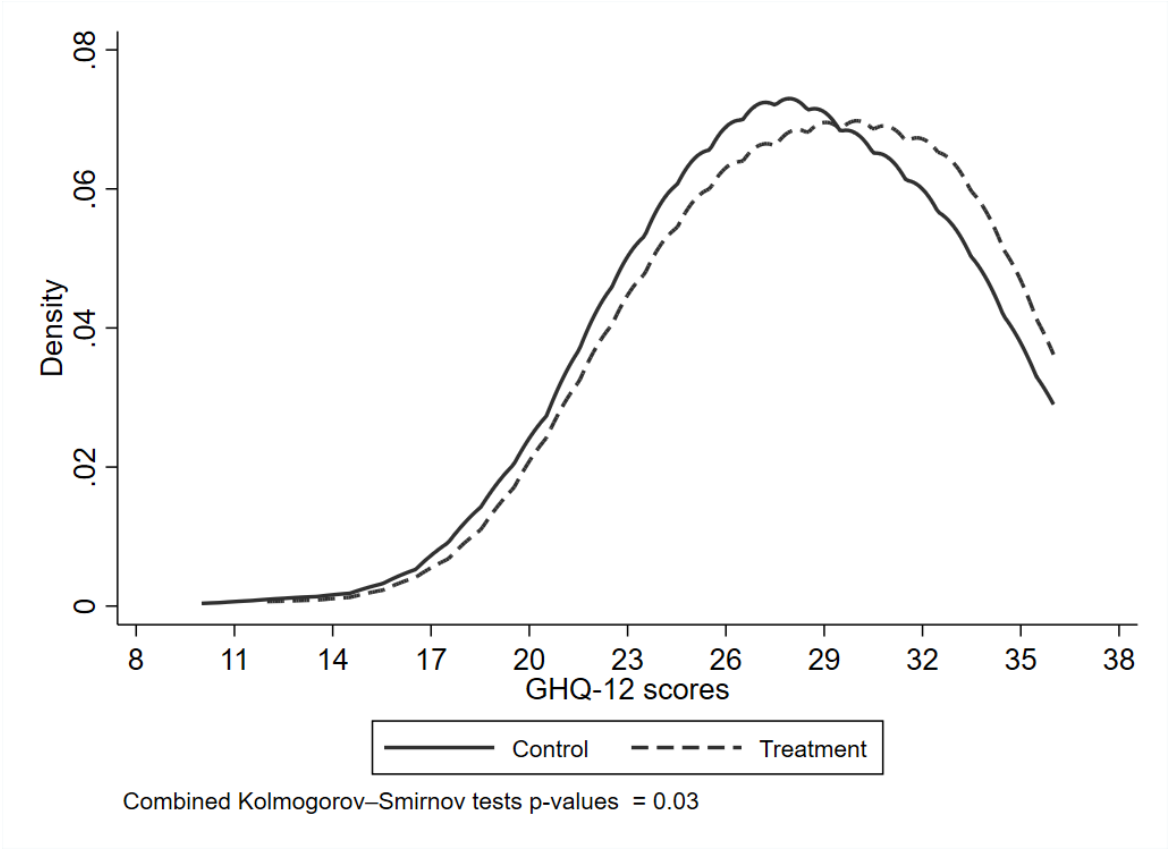
Note. Panel A shows the level of parental mental health (K-6 score) throughout their children's developmental periods. Panel B shows the level of depressed parents throughout their children's developmental periods. Both measures partialled out the age and age-squared. Data source: China Family Panel Study 2010, number of sample parents (aged 20-60): 21,549 observations.

Figure 2: Timeline of the Experiment



Note. This figure shows the timeline of the experiment. The intervention lasted for 4 months, starting from the 1st of March till the 30th of June in 2021. We list the monthly theme of the program during the treatment period.

Figure 3: Distribution of GHQ-12 Across Treatment and Control Groups



Note. This figure illustrates the distribution of GHQ-12 scores across the treatment and control groups. Higher GHQ scores indicate better mental health. The P-value for the combined Kolmogorov–Smirnov test is 0.03.

Table 1: Balance Test

	(1) Control	(2) Mean difference (T-C)
Panel A. Parent's characteristics		
Age	42.597	-0.228
	(4.735)	(0.218)
Mother/not	0.803	0.006
	(0.398)	(0.024)
Urban Hukou	0.295	-0.013
	(0.456)	(0.031)
Migrant	0.041	-0.003
	(0.199)	(0.013)
Married	0.933	0.017
	(0.251)	(0.013)
Income (< $15K)	0.242	-0.013
	(0.428)	(0.025)
Income ($15k-30k)	0.397	0.014
	(0.490)	(0.026)
Income ($30k-60k)	0.197	0.005
	(0.398)	(0.017)
Income (> $60k)	0.164	-0.007
	(0.370)	(0.029)
Panel B. Children's characteristics		
Age	14.476	-0.040
	(0.578)	(0.139)
Male	0.521	0.023
	(0.500)	(0.019)
Urban Hukou	0.474	-0.018
	(0.500)	(0.032)
No siblings	0.284	0.019
	(0.451)	(0.022)
Height	161.784	0.037
	(7.648)	(0.736)
Weight	50.745	-0.811
	(10.426)	(0.699)
Stress score	13.237	-0.084
	(4.142)	(0.297)
CES-D 10	8.719	-0.586
	(5.578)	(0.360)

Note. This table shows the regression results attempting to verify the randomization of the classroom assignments. Panel A reports results on parents' characteristics, while Panel B reports results on students' characteristics. Column 1 reports the summary statistics for the control sample. Column 2 reports the differences in means for each variable between treatment and control groups, as well as the standard errors for item-wise regression using the variables labeled in the first column as the dependent variables. Classroom-level clustered standard errors are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

Table 2: Program Impacts on Parental Mental Health

	(1) Control mean	(2) ITT	(3) Permutation	(4) WCB	(5) TOT
Panel A. Mental Health					
GHQ (Likert)	28.006	0.792***	0.009	0.011	1.891***
	(4.690)	(0.289)			(0.707)
GHQ (Standardized)	0.000	0.169***	0.009	0.011	0.403***
	(1.000)	(0.062)			(0.151)
Very happy last week	0.423	0.060**	0.043	0.038	0.170**
	(0.494)	(0.027)			(0.080)
N	848	1,852			1,852
Panel B. Three dimensions					
Social dysfunction	0.000	0.152**	0.002	0.016	0.363**
	(1.000)	(0.059)			(0.143)
Anxiety	0.000	0.156**	0.002	0.019	0.372**
	(1.000)	(0.064)			(0.157)
Loss of confidence	0.000	0.118**	0.011	0.045	0.281**
	(1.000)	(0.053)			(0.129)
N	848	1,852			1,852

Note. This table shows ITT estimates of equation (1) for parental mental health. We controlled for strata fixed effects to take into account of stratified randomization design. Column 1 reports the means and the standard deviations for the corresponding outcome variables for those in control classes. Column 2 reports the ITT estimates and standard errors, while Columns 3 and 4 report the associated Permutation P-value after 2000 stratified clustered resamplings and Wild Cluster Bootstrap P-value after 9999 resamplings. Column 5 reports the TOT estimates of equation (2). First-stage Kleibergen-Paap rk Wald F statistic = 108.624. Classroom-level clustered standard errors are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

GHQ score *adds up all 12 items and ranges between 0 and 36.* ***Feel very happy*** *is a dummy variable and measures parents' self-evaluated happiness in the past week. In Panel B, We use an inverted standardized score for* ***Social dysfunction***, ***Anxiety***, *and* ***Loss of confidence*** *so that a higher score indicates a better mental health status in that dimension. For comparison purposes, we constructed standardized values relative to the control group.*

Table 3: Mechanisms: Parenting skills, time investments, and Child Abilities

	(1) Control mean	(2) ITT	(3) Permutation	(4) WCB	(5) RW p-values	(6) TOT
Panel A: Indices						
Overall Index	0.000	0.313***	0.000	0.000		0.747***
	(1.000)	(0.097)				(0.229)
Parenting skill	0.000	0.197***	0.000	0.002	0.013	0.470***
	(1.000)	(0.058)				(0.146)
Time investment	0.000	0.154**	0.020	0.029	0.058	0.368**
	(1.000)	(0.067)				(0.156)
Child mental health	0.000	0.230**	0.004	0.016	0.058	0.548**
	(1.000)	(0.093)				(0.217)
Child non-cognitive ability	0.000	0.172**	0.004	0.027	0.058	0.411**
	(1.000)	(0.075)				(0.168)
N	848	1,852				1,852
Panel B: Parent outcomes						
Parenting skills:						
Empathy	0.000	0.100**	0.052	0.042	0.050	0.240**
	(1.000)	(0.046)				(0.110)
Democratic parenting	0.789	0.039**	0.029	0.030	0.050	0.094**
	(0.408)	(0.017)				(0.043)
Understand child's feeling	2.380	0.135**	0.026	0.031	0.050	0.322**
	(0.951)	(0.060)				(0.146)
Encourage child's hard work	2.264	0.172***	0.003	0.003	0.020	0.410***
	(0.994)	(0.057)				(0.143)
Time inputs:						
Time investment weekday	3.725	0.513**	0.010	0.020	0.027	1.225**
	(3.288)	(0.204)				(0.483)
Time investment weekend	5.413	0.408*	0.078	0.097	0.061	0.974*
	(3.649)	(0.224)				(0.532)
N	848	1,852				1,852
Panel C: Child outcomes						
Mental health:						
Feel happy	4.890	0.256**	0.024	0.035	0.073	0.610**
	(5.771)	(0.114)				(0.261)
Depressed (CES-D)	0.364	-0.048*	0.087	0.093	0.073	-0.114*
	(0.482)	(0.026)				(0.060)
Less stress	0.000	0.206**	0.016	0.034	0.073	0.491**
	(1.000)	(0.091)				(0.216)
Non-cognitive ability:						
Empathy	0.000	0.120*	0.105	0.115	0.073	0.286*
	(1.000)	(0.071)				(0.156)
Positive personality	0.000	0.169**	0.008	0.012	0.059	0.403***
	(1.000)	(0.067)				(0.155)
N	848	1,852				1,852

Note. This table shows ITT estimates for the three mechanisms - parenting skills, time investments, child mental health, and child non-cognitive skills - using equation (1). We controlled for strata fixed effects to take into account of stratified randomization design. Column 1 reports the means and the standard deviations for the control group. Column 2 reports ITT estimates and standard errors. Columns 3 and 4 report the associated Permutation P-value after 2000 stratified clustered resamplings and Wild Cluster Bootstrap P-value after 9999 resamplings. Column 5 reports Romano and Wolf (2005) adjusted p-values as discussed in Section 4.2. Lastly, Column 6 reports the TOT estimates. First-stage Kleibergen-Paap rk Wald F statistic: 108.624. Classroom-level clustered standard errors are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

Table 4: Cost-effectiveness Analysis

Panel A: Benefits	
1. Total benefit per parent	0.17 SD in mental health
2. Total benefit per class	1.4 depression incident averted
Panel B: Organizational costs (include in-kind)	
Total organizational cost	$1,400
Organizational cost per class	**$30**
Organizational cost per parent	**$0.64**
Cost-benefit ratios:	
Cost per 0.1 SD improvement in mental health	**$17.65**
Cost per depression incident averted	**$21.43**
Panel C: Opportunity costs	
Cost per class (account for parental time input)	$1,170
Cost per class (include parents' opportunity costs)	$1,200
Cost per parent (include parents' opportunity costs)	**$25.53**
Cost per depression incident averted (include opportunity costs)	$857

*Note. This table reports the cost-effectiveness analysis of our program. Panel A reports the average benefit per parental mental health improvement and the average benefit – depression averted – per class. We use the conservative ITT estimates for the depression indicator – 0.03pp and the unit of depression averted per class = 0.03 * 47 ≈ 1.4. Panel B reports the organizational costs, while Panel C reports the costs taking into account potential opportunity costs of parental time investments. All costs are converted into USD using the exchange rate of 7.2 RMB per USD. Total organizational cost is $1,400 for 48 classes. Organizational cost per class = $1,400/48 = $30. Considering the average class size is 47, organizational cost per parent = $30/47 ≈ $0.64. The cost-benefit ratio (cost per benefit) is calculated by dividing the total cost by the total benefit. Cost per 0.1 SD improvement in mental health = $30/0.17 * 0.1 ≈ $17.65. Cost per depression incident averted = $30/1.4≈$21.43. In terms of opportunity costs, the hourly income is 1.5 × the local minimum wage (20 RMB). Total opportunity cost per class= hourly income × monthly input × number of compliers in each month= 30×4×(20×3+10)=8,400 RMB ($1,170). Cost per class including parents' opportunity costs = $1,170+$30 = $1,200. Cost per parent accounting for parents' opportunity costs = $1,200/47 ≈ $25.53. Cost per depression incident averted accounting for parents' opportunity costs = $1,200/1.4 ≈ $857.*

ONLINE APPENDIX

Supplementary material to *Helping Parents Combat Middle-School Blues: Evidence from a Cluster Randomized Controlled Trial on Empathy and Parental Involvement*

Qinyou Hu, and Yiming Xia

A Details in Questionnaire

A.1 Parents' Outcome Measures

We list the detailed questions and measures in the parent survey below:

- Participation in empathy-related activities: The variables measure the take-up of the program. We asked parents whether they ever watched movies or read short articles on empathy with their kids at least once or at least once per month in the past semester.
- Parental mental health status: We use the 12-item General Health Question (GHQ-12) to elicit mental health (Chan, 1985).
- Parenting style: We provide detailed explanations of the four types of parenting styles - authoritative, authoritarian, permissive, and neglecting, and ask the parents to select one of the types that is the most applicable to them.
- Parental responsiveness: Two items reported by students: "Can your parents tell your feelings and emotional changes?" and "Will they encourage you when they see your hard work?"
- Parents' empathy: Two items on perspective taking and empathetic concerns, and they are constructed following the same method as constructing students' outcome measure.

- Parental time investment: The frequency of parent-child activities per day, including reading, checking homework, playing, and conducting general education with kids on weekdays and weekends, respectively, over the past week.

- Parental monetary investment: It is sensitive to directly inquire about their monetary investment in children. We instead only ask them about the investment as a proportion of their total income. We divided it into five categories: 5% of the total income, 5% - 10%, 10% - 25%, 25% - 50%, and more than 50% of the total income.

- Parental attitudes towards monetary investment (after school tutoring): Parents are asked to select whether they will send their kids to after-school tutoring in three hypothetical settings: (i) when their best friends' children were sent to after-school tutoring; (ii) when the best students in the class were sent to after school tutoring; and (iii) when most of the students in the class were sent to after school tutoring. We also elicit the perceived value of cram schools by asking parents to score whether the after-school tutoring is good for students' test scores and whether the after-school tutoring is good for students' mental health for a hypothetical struggling student, with a scale of 1-100.

A.2 Students' Outcome Measures

In the analysis of mechanisms, we use students' mental health conditions, which are composed of CES-D scores, happiness, and stress resistance, as well as non-cognitive abilities, which encompass empathy and positive personality traits. The detailed measures and questions from the student survey are listed below.:

- Depressed: Mental health is measured using the 10-item Center for the Epidemiological Studies of Depression Short Form, or CES-D-10, is a 10-item Likert scale questionnaire (Yang et al., 2018).[1] The depression indicator is generated with a threshold

[1]The items are also employed in China Family Panel Studies (CFPS) 2012 survey.

value of 12. The *inverse CES-D index* is constructed by 30 minus the CES-D score so that a higher score indicates better mental health status.

- Happiness: We construct happiness scores using a scale of 1-7, with 7 being the happiest.

- Stress score: We elicit students' stress by three categories of sources: (i) studies at school, (ii) peer relationships, and (iii) rank/test scores in the class. For each item, we use a 7-point Likert scale for both baseline and follow-up surveys: from the least stressed (1) to the most stressed (7). We then construct the *inverse stress index* consisting of these three components so that a higher score indicates less stress.

- Empathy: To avoid long questionnaires, we use a 9-item empathy measurement to explore two dimensions of empathetic concerns and perspective taking, which is also used in Alan et al. (2021). For most items, we use a 7-point Likert scale for both baseline and follow-up surveys: from completely disagree (1) to completely agree (7). We added another dimension, fantasy, to the follow-up survey. The questions include the hypothetic scenarios about helping other children in difficulty, doing others a favor, helping my mother do housework, becoming a charitable person, and rescuing a drowning child adapted from the official guide from Centers for Disease Control (CDC) (Dahlberg et al., 2005). For each scenario, we asked students whether they had ever imagined the scenarios, and they were asked to choose (1) Never, (2) Sometimes, or (3) Very frequent.

- Positive personality trait: four aspects of self-image were measured by four single-item questions: (i) I am satisfied with myself (self-satisfied); (ii) I have many valuable traits (self-worth); (iii) I can do well in most cases (self-confident); (iv) I am not worse than others and proud of myself (self-esteem). For each item, we use a 7-point Likert scale for both baseline and follow-up surveys: from completely disagree (1) to completely agree (7). In the empirical analysis, we use inverse covariance

matrix weighting methods to construct the self-esteem index that includes these four components. Additionally, we measure students' perseverance by asking students whether they agree or disagree with the statement "Frustration and difficulty won't stop me from reaching my goals." We use a 7-point Likert scale for both baseline and follow-up surveys: from completely disagree (1) to completely agree (7).

- Time with parents: As a cross-check of parental time investment, we ask students to count the total number (ranges from 0 - 7) of each activity that has parents involved in a normal week in the past semester. The activities include having dinner, talking/discussing school life, watching TV, checking homework, and playing outdoor activities.

- Parental responsiveness: Students are asked (i) whether their parents will praise them for what they did? and (ii) when they perform well, whether their parents will notice and let them know? For each question, we ask students to choose from never or seldom (1), sometimes (2), often (3), and always (4).

B Additional Analysis

B.1 Impact Heterogeneity on Happiness

We also analyze the effect heterogeneity on parents' happiness – the dummy variable of "feel very happy" – in Appendix Figure C6. We find that ITT effects are mainly concentrated among those parents who are middle-income, younger, with only one child, as well as those whose children have better baseline mental health. Interestingly but not statistically significant, we find that mothers, compared to fathers, are significantly more likely to feel very happy after the intervention, while the patterns of ITTs on GHQ are opposite.

B.2 Time Investments

We also investigate the treatment effects on detailed time use reported by the students as a robustness check for parental time investment results in Table 3. Results are shown in Appendix Table C11. Consistent with findings from parents' responses, the program significantly leads to students in the treated groups being more likely to talk with parents and to have homework checked by parents. Treated students also reported a significantly higher frequency of having outdoor activities with their parents.

B.3 Monetary Investments

To test the crowding-out effect of time investments, we report the effects on monetary investments and parents' beliefs on one of the primary education expenditures – after-school tutoring in Panel B in Appendix Table C10. In nowadays China, more than 60% of parents spend zero hours accompanying their kids on a typical weekday. At the same time, over 90% of them send their children to attend after-school tutoring classes, the so-called "cram schools."[2] From the survey responses, about 27% parents in our study sample point out it is due to that other children's parents also send the kids to the classes, the fear for legging behind other students given the fierceness of competition at school, and the mentality of feeling obliged to do "whatever everyone else is doing." Another 11% point out that they send out their kids because they do not have time to accompany the children. Thus, we would like to test whether our directed parental involvement program may affect parents' attitudes toward their private tutoring investment decisions. The monetary investment is measured as the percentage of total income. From the estimates shown in Column 2, it appears that the program did not crowd out the monetary investment, and it shows precisely no effect on parents' beliefs in cram schools. The probability of deciding to send their children to cram

[2] http://www.chinadaily.com.cn/a/201806/15/WS5b2300a5a310010f8f59d147_1.html.

schools is higher in scenarios 2 and 3, where for scenario 2, the hypothetical case is that the best student in the class is taking the private tutoring classes, and for scenario 3, most (more than 50%) students in the class are seeking for extra tutorials.

References

Alan, S., C. Baysan, M. Gumren, and E. Kubilay (2021). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *The Quarterly Journal of Economics 136*(4), 2147–2194.

Chan, D. W. (1985). The chinese version of the general health questionnaire: does language make a difference? *Psychological Medicine 15*(1), 147–155.

Christensen, M. K., C. C. W. Lim, S. Saha, O. Plana-Ripoll, D. Cannon, F. Presley, N. Weye, N. C. Momen, H. A. Whiteford, K. M. Iburg, and et al. (2020). The cost of mental disorders: a systematic review. *Epidemiology and Psychiatric Sciences 29*, e161.

Dahlberg, L. L., S. B. Toal, M. H. Swahn, and C. B. Behrens (2005). Measuring violence-related attitudes, beliefs, behaviors, and influences among youths: a compendium of assessment tools. Technical report, Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Delgadillo, J., S. Ali, K. Fleck, C. Agnew, A. Southgate, L. Parkhouse, Z. D. Cohen, R. J. DeRubeis, and M. Barkham (2022). Stratified care vs stepped care for depression: A cluster randomized clinical trial. *JAMA Psychiatry 79*(2), 101–108.

Ding, R., D. Zhu, Y. Wang, M. Yong, X. Shi, and P. He (2022). Medical service utilisation and direct medical cost of depression: a cross-sectional analysis of urban medical claims data from china. *BMJ Open 12*(9), e056422.

Fuhr, D. C., B. Weobong, A. Lazarus, F. Vanobberghen, H. A. Weiss, D. R. Singla, H. Tabana, E. Afonso, A. De Sa, E. D'Souza, et al. (2019). Delivering the thinking healthy programme for perinatal depression through peers: an individually randomised controlled trial in india. *The Lancet Psychiatry 6*(2), 115–127.

Greenberg, P. E., A.-A. Fournier, T. Sisitsky, M. Simes, R. Berman, S. H. Koenigsberg, and R. C. Kessler (2021). The economic burden of adults with major depressive disorder in the united states (2010 and 2018). *Pharmacoeconomics 39*(6), 653–665.

Huang, R., C. Yan, Y. Tian, B. Lei, D. Yang, D. Liu, and J. Lei (2020). Effectiveness of peer support intervention on perinatal depression: A systematic review and meta-analysis. *Journal of Affective Disorders 276*, 788–796.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Richards, D., A. Enrique, N. Eilert, M. Franklin, J. Palacios, D. Duffy, C. Earley, J. Chapman, G. Jell, S. Sollesse, et al. (2020). A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. *NPJ Digital Medicine 3*(1), 85.

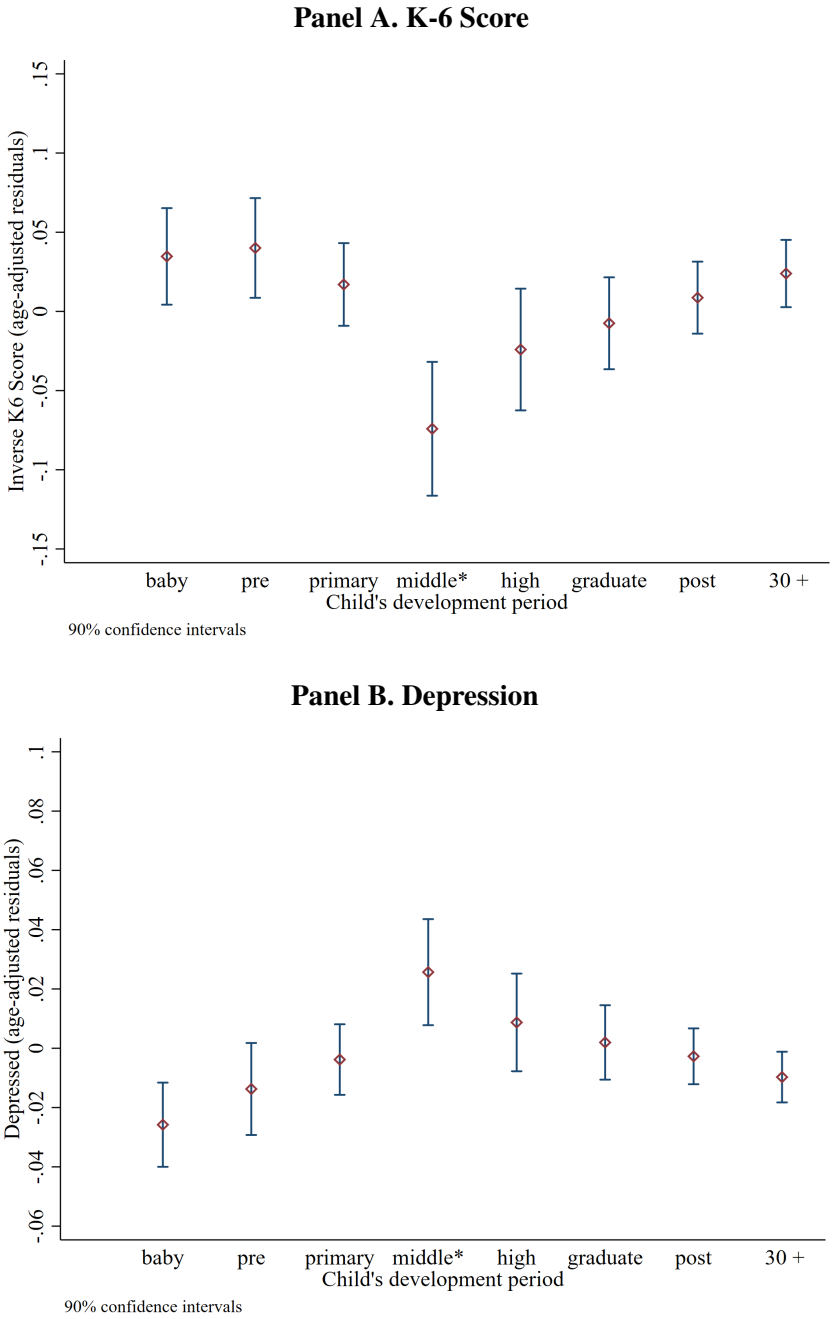
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Xu, J., J. Wang, A. Wimo, and C. Qiu (2016). The economic burden of mental disorders in china, 2005–2013: implications for health policy. *BMC Psychiatry 16*, 1–9.

Yang, W., G. Xiong, L. E. Garrido, J. X. Zhang, M.-C. Wang, and C. Wang (2018). Factor structure and criterion validity across the full scale and ten short forms of the ces-d among chinese adolescents. *Psychological Assessment 30*(9), 1186.

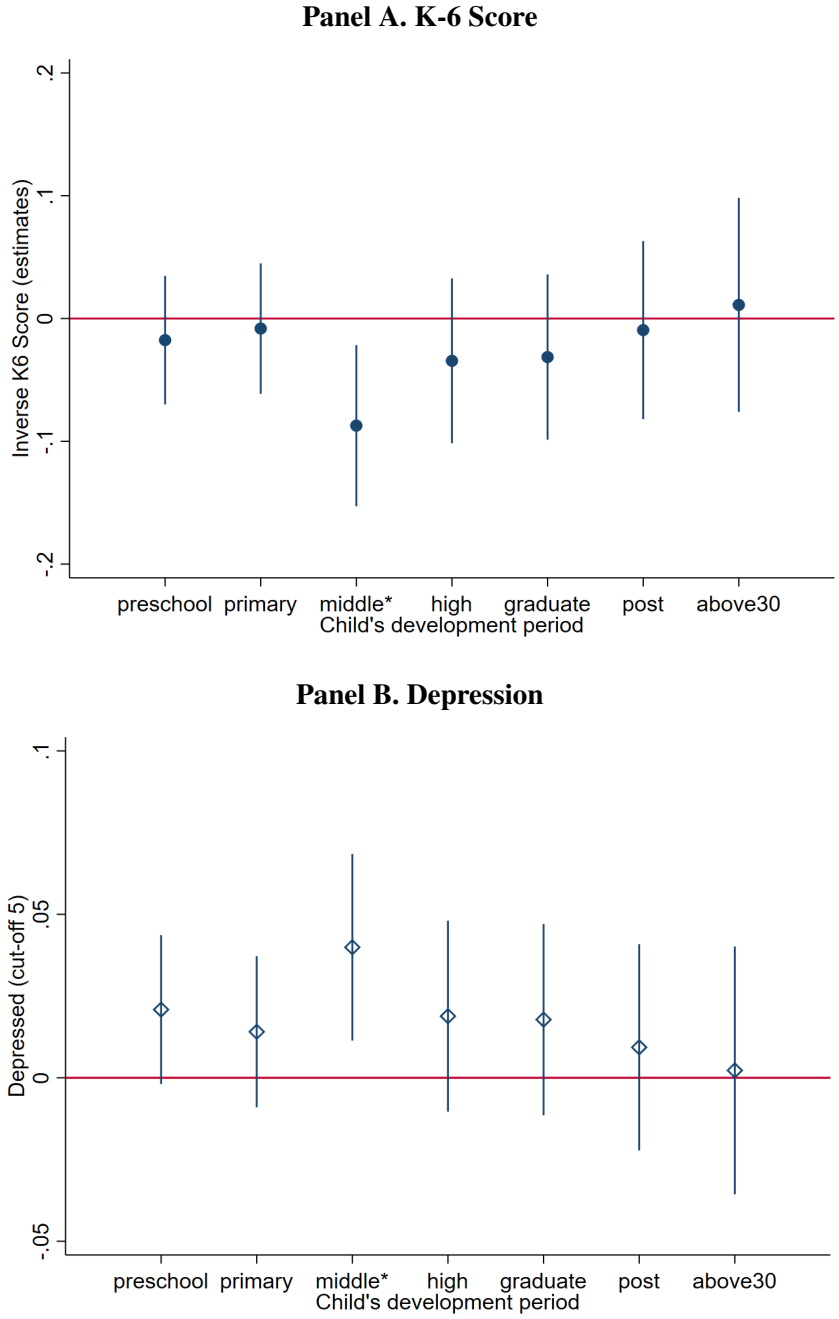
C Tables and Figures

Figure C1: Parental Mental Health Across Children's Developmental Periods



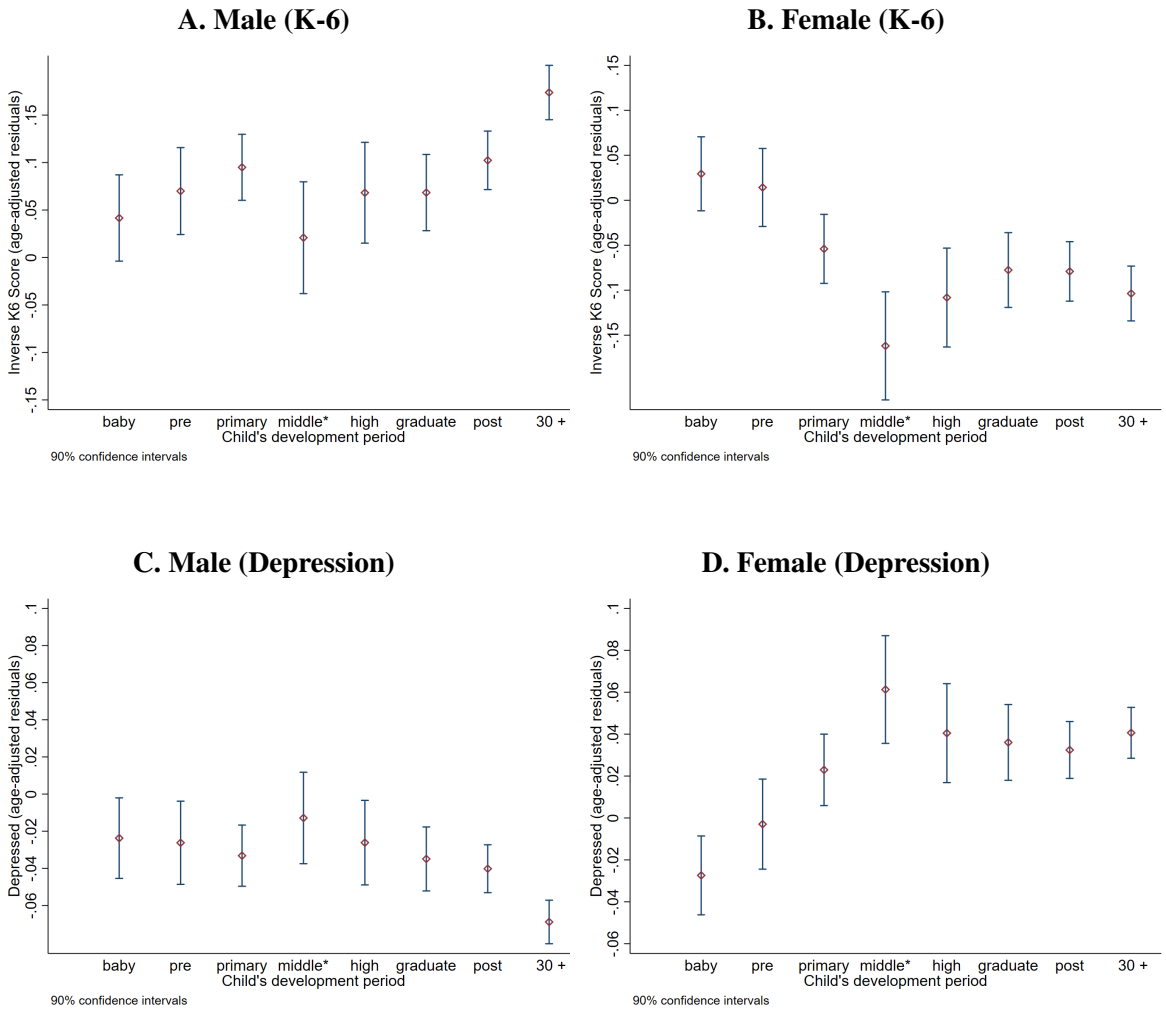
Note. Panel A shows the level of parental mental health (K-6 score) throughout their children's developmental periods. Panel B shows the level of depressed parents throughout their children's developmental periods. Both measures partialled out the age and age-squared. We report the 90% confidence intervals around the means. Data source: China Family Panel Study 2010, number of sample parents (aged 20-60): 21,549 observations.

Figure C2: Parental Mental Health Across Children's Developmental Periods, Relative to Infant Period



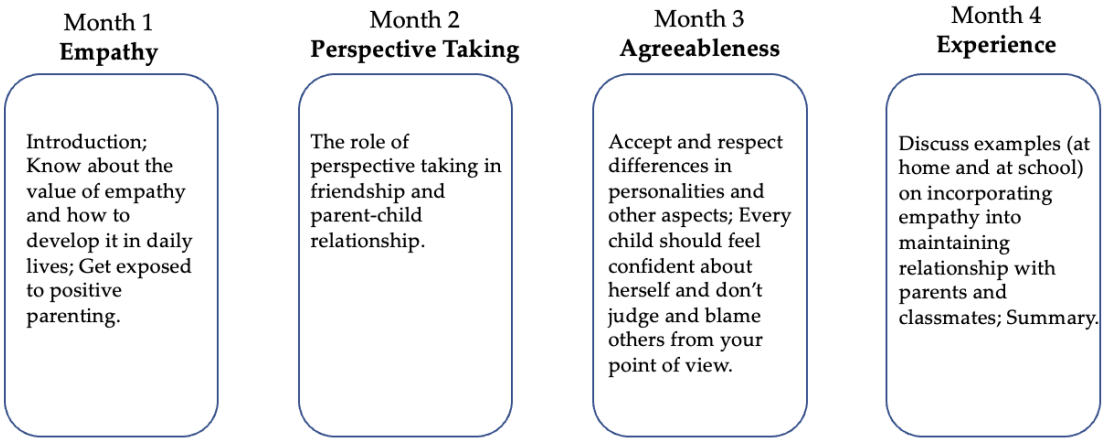
Note. Panels A and B show the regression estimates describing the difference in parental mental health (K-6) across child developmental periods against the reference group (parents with children aged 0-2). The regression is estimated on adults between 20 and 60 years old and controls for age, age squared, the number of children, income, hukou type (agricultural), urban residents, and education level. We report the 90% confidence intervals around the means. Data sources: China Family Panel Study 2010, 21,549 observations.

Figure C3: Middle-School Blues by Gender



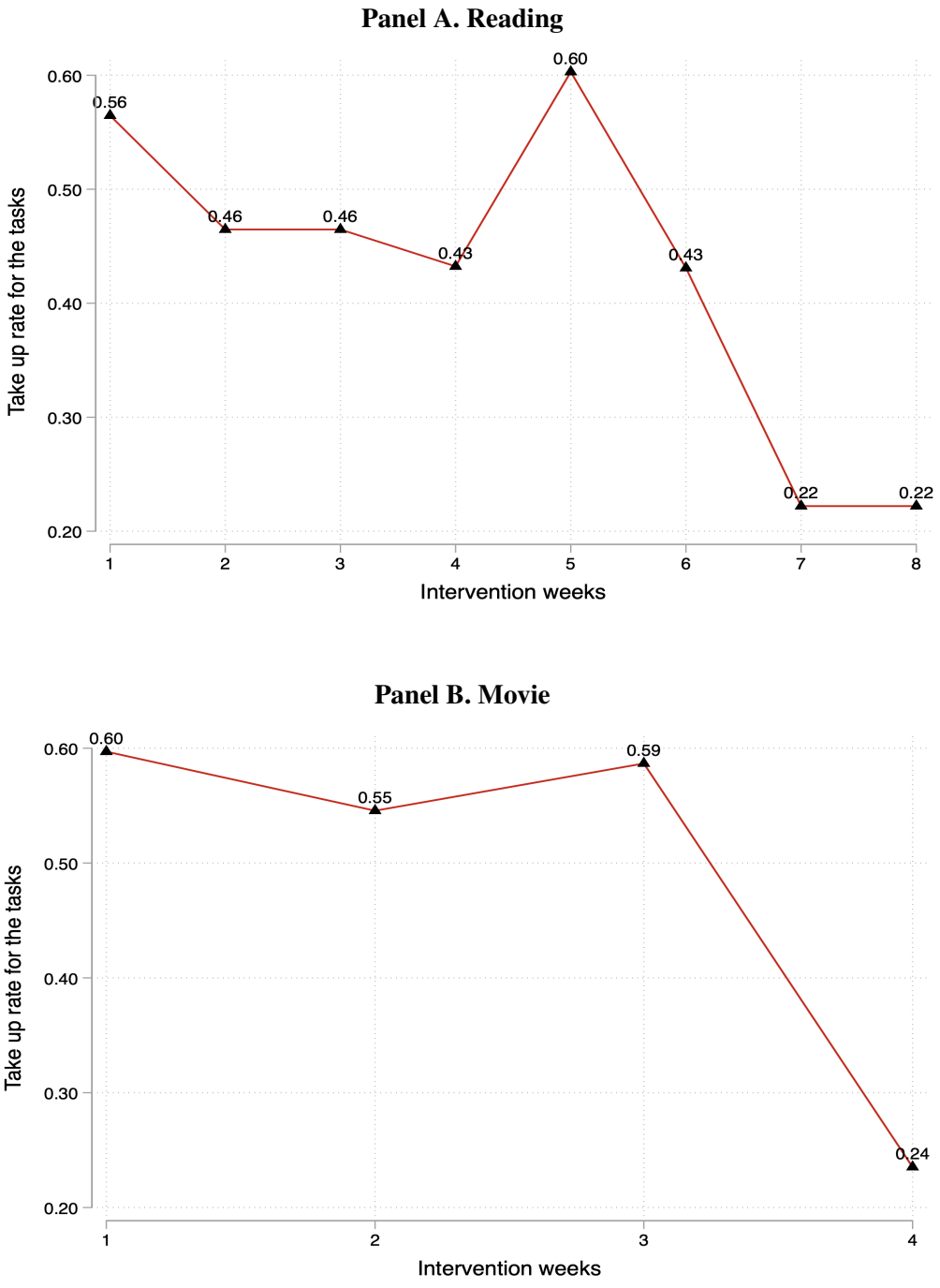
Note. Panels A and B show the levels of fathers' and mothers' adjusted K-6 scores across their children's developmental periods, respectively. Panels C and D show the same patterns for depression. Both measures partialled out the age and age-squared. We report the 90% confidence intervals around the means. Data sources: China Family Panel Study 2010, 21,549 observations.

Figure C4: Theme of the Program



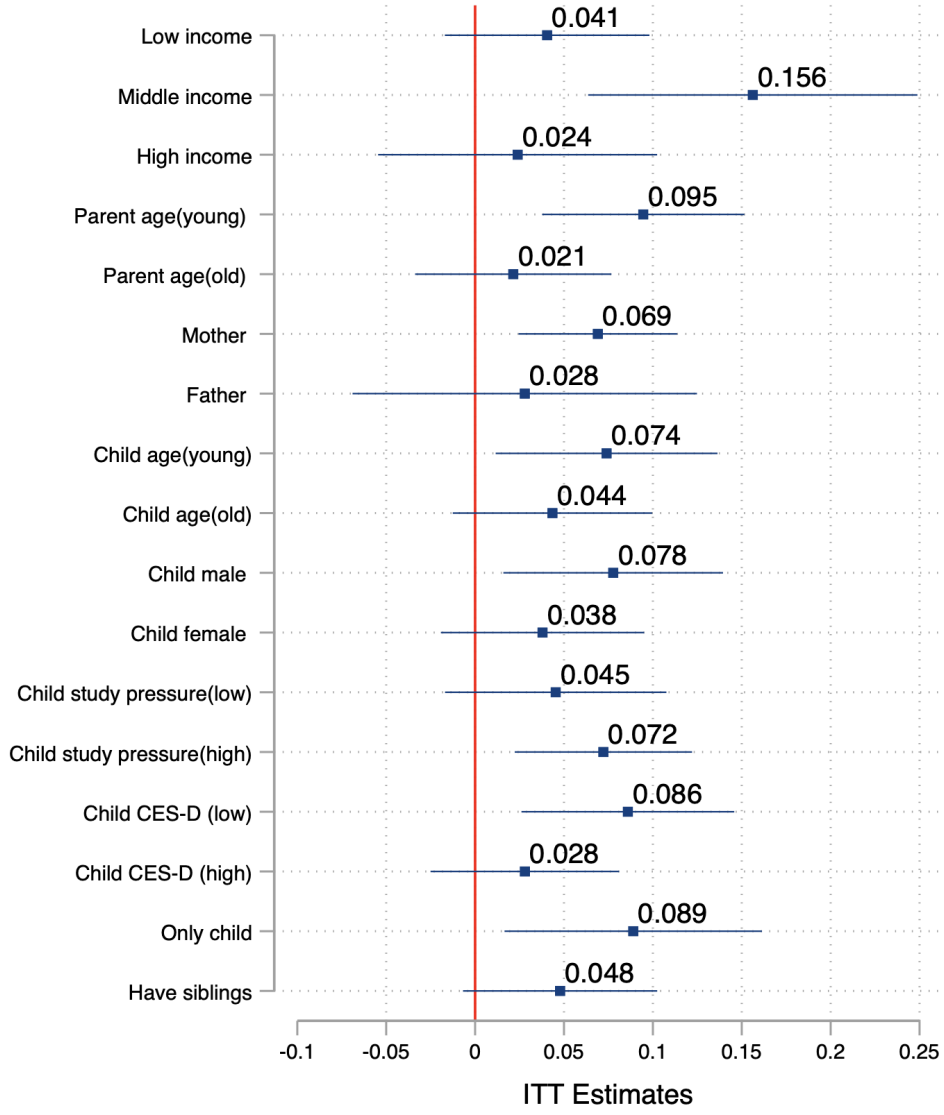
Note. The figure details the theme of our 4-month parental involvement program. We rely on two books by A.Ciaramicoli and K. Ketcham as a reference for these themes and materials. We expand the details of the program in Table C2.

Figure C5: Detailed Task Completion Rates



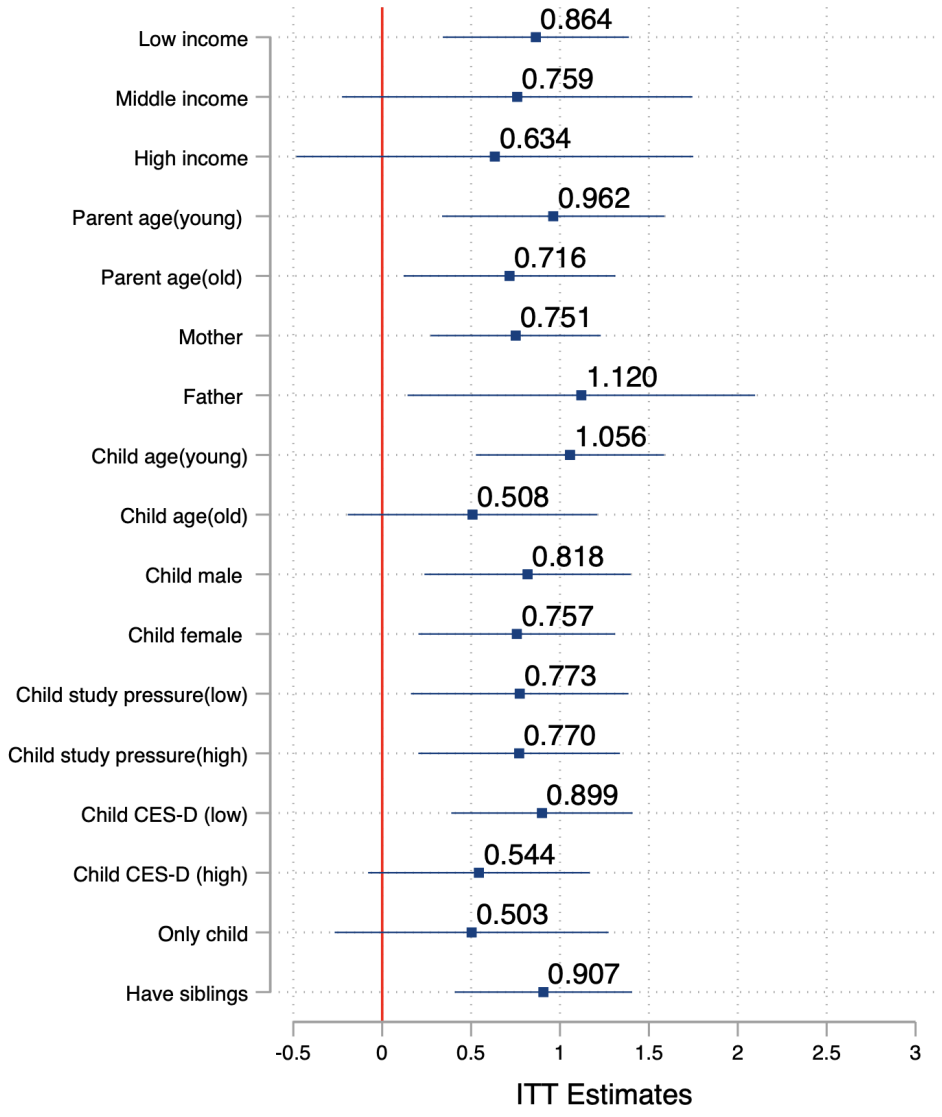
Note: Panel A shows the completion rate of the biweekly reading activities. Panel B shows the completion rate of the monthly movie activities. The numbers are calculated by the total number of those who completed the specific task divided by the eligible parents (i.e., those in the treatment group, N = 1,004).

Figure C6: Intention to Treat Estimates for Happiness, Heterogeneous Effects



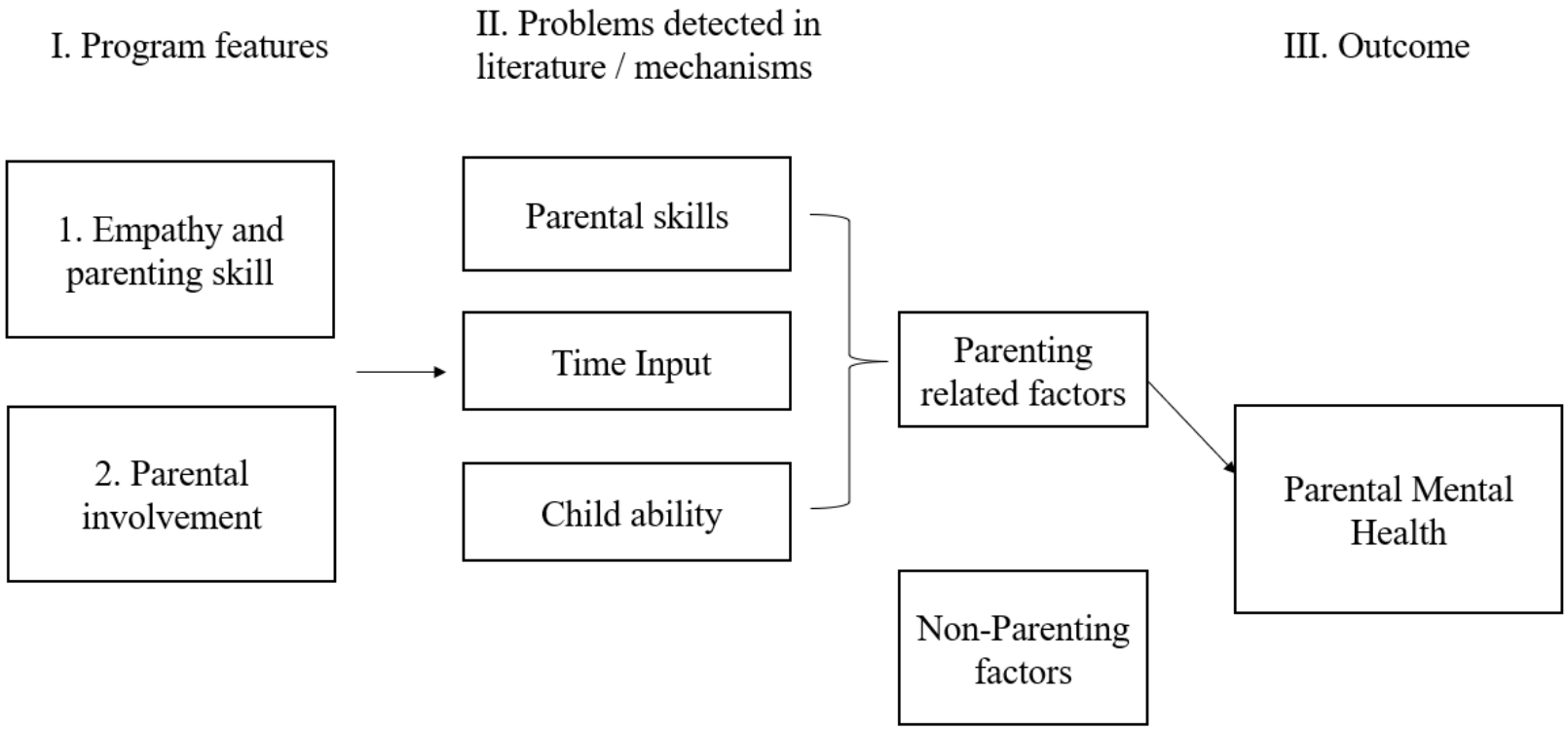
Note. This figure illustrates the heterogeneous treatment effects for "feel very happy" using subgroup analysis. "Feel very happy" is a dummy variable of happiness from the survey question "Do you feel happy in the past week?" It takes the value one if parents answer "Feel very happy" and zero otherwise. Low income refers to families with an annual income lower than $30k. Middle income refers to families with an annual income ranging from $30k to $60k. High income refers to families with an annual income higher than $60k. The splitting cut-off values for other continuous variables are the medians of parents' age, baseline child age, child pressure score, and child CES-D score of the study sample. The 90% confidence intervals are computed using the classroom-level clustered standard errors.

Figure C7: Intention to Treat Estimates for Parental Mental Health (GHQ-12 Likert), Heterogeneous Effects



Note. This figure illustrates the heterogeneous treatment effects for the GHQ-12 score using subgroup analysis. Low income refers to families with an annual income lower than $30k. Middle income refers to families with an annual income ranging from $30k to $60k. High income refers to families with an annual income higher than $60k. The splitting cut-off values for other continuous variables are the medians of parents' age, baseline child age, child pressure score, and child CES-D score of the study sample. The 90% confidence intervals are computed using the classroom-level clustered standard errors. We report the Romano-Wolf adjusted p-values for the estimates in Appendix Table C8.

Figure C8: Potential Mechanisms



Note. This figure shows the logic behind our experiment's effectiveness in promoting parental mental health. The intervention has two main features which target the mechanisms discussed in the literature. The intervention improves three main parenting-related factors, ultimately leading to improved parental mental health.

Table C1: Middle-School Blues in China

	(1) OLS	(2) OLS	(3) OLS
Middle school child	-0.185***	-0.133**	-0.100*
	(0.059)	(0.058)	(0.059)
Age	Yes	Yes	Yes
Demographics		Yes	Yes
Province FE			Yes
N	3,246	3,214	3,214
R-squared	0.0086	0.051	0.076

Note. This table describes the relationship between parental mental health (Inverted K-6) and having middle school children. It examines the existence of middle school blues using the nationally representative sample from the China Family Panel Study 2010 (CFPS). Column 1 reports the OLS estimates after controlling age and age-squared. Columns 2 and 3 report the OLS estimates, which account for demographics (including the number of children) and the province fixed effects sequentially. Our preferred estimates are those reported in Columns 2 and 3. The sample is restricted to parents (ages 20-60) who have middle-school-aged children (ages 12-14) and those with relatively younger children (ages 11). Standard errors clustered at the community level are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

Table C2: Intervention Contents

Time	Tasks
Week 1 (M1)	**Reading task**: read a short article on (i) What is empathy? and (ii) The importance and value of empathy. **Movie of the 1st month**: watch "Looking Up" together, then discuss the parenting styles in the movie with your child
Week 3	**Cases and examples**: read a short article on (i) Parents incorporate empathy into parenting styles and (ii) Positive parenting skills
Week 5 (M2)	**Reading task**: read a short article on (i) Perspective taking and (ii) The importance of perspective taking on friendship and parent-child relationship **Movie of the 2nd month**: watch "Wonder" together, then think about the script "Auggie can't change the way he looks. But maybe we can change how we look at him."
Week 7	**Cases and examples**: read a short article on (i) Self-centeredness and (ii) How to become less self-centered
Week 9 (M3)	**Reading task**: read a short article on (i) Personality and multiple intelligences and (ii) The importance of being unique and respecting each other **Movie of the 3rd month**: watch "Taare Zameen Par" together, then focus on discussing the script "Every child is like a shining star, we should discover the uniqueness of each child from different perspectives." with your child
Week 11	**Cases and examples**: read a short article on (i) How to educate your child according to their individual uniqueness and (ii) How to teach your child to embrace others' uniqueness, especially when they look different.
Week 13 (M4)	**Reading task**: read a short article on (i) Lack of empathy and peer relationship and (ii) How parents can help children to get through poor peer relationship **Movie of the 4th month**: watch "Better Days" together, then focus on discussing the script "why we can't learn sympathy until becoming an adult?" with your child
Week 15	**Cases and examples**: read a short article on (i) Emotional skills help students improve peer relationships and (ii) Lack of empathy fosters cold and distant relationships with peers, creating more adverse consequences.

Note. This table shows the detailed contents of the intervention. Parents are encouraged to discuss and exchange views on the task content with their children and submit a short reflection essay to the platform once they finish the task. The first column shows the time when the tasks were sent. Each task was delivered via the WeChat *group of treated classes by the class teacher on Friday evenings. The second column summarizes the main components of each task. The short articles on the biweekly reading tasks were uploaded to the platform that we created one day prior to the delivery date. The estimated time for the reading task is about 30-45 min. The monthly activities (watching movies) were assigned and announced on the platform on the first Friday of each month. The estimated time for the movie task is about 90-120 min.*

Table C3: Baseline Parental Involvement

	(1) Non-engagement	(2) Frequency
Eat	0.059	3.191
	(0.236)	(2.463)
Talk	0.096	3.926
	(0.294)	(2.740)
Watch TV	0.565	1.107
	(0.496)	(1.848)
Check homework	0.398	2.439
	(0.490)	(2.835)
Outdoor activities	0.418	1.615
	(0.493)	(2.096)
N	2,246	2,246

Note. This table reports the descriptive statistics for baseline parental involvement in the study sample. Column (1) reports the means and standard deviations for indicators of five activities measuring parental involvement. Each variable takes value 1 if it occurred at least once in the past week. Column (2) reports the means and standard deviations for the average number of the activities in the past week. Standard deviations are presented in parentheses.

Table C4: Attrition in Parent Survey

Panel A. Attrition rate		
	(1)	(2)
Fraction of nonresponding parents	Control	T-C
	0.174	-0.001
	(0.379)	(0.023)
Panel B. Testing selective attrition		
	Attrition	Attrition * Treat
Age	0.016	0.010
	(0.043)	(0.061)
Male	0.037	-0.089*
	(0.034)	(0.052)
Urban hukou	0.004	-0.052
	(0.046)	(0.054)
Only child	0.055	-0.061
	(0.046)	(0.056)
Height in cm	0.178	0.362
	(0.684)	(0.862)
Weight in half kilo	0.887	1.424
	(1.810)	(2.403)
Bullying perpetrator	0.013	0.012
	(0.035)	(0.055)
Bullying victim	-0.009	0.054
	(0.034)	(0.050)
Number of friends	-0.317***	0.088
	(0.103)	(0.140)
Member of exclusive group	0.001	-0.055
	(0.037)	(0.054)
Empathy score	-1.242	0.391
	(0.784)	(1.005)
Self-satisfied	-0.453***	0.024
	(0.112)	(0.147)
Self-worth	-0.493***	0.269
	(0.116)	(0.164)
Self-confident	-0.214	-0.296
	(0.159)	(0.186)
Self-esteem	-0.301	-0.041
	(0.211)	(0.241)
Perseverance	-0.236	-0.074
	(0.158)	(0.200)
Stress score score	0.292	-0.132
	(0.326)	(0.456)
CESD 10-item	1.133**	0.306
	(0.540)	(0.587)
Weekly interaction with parents	-0.773**	-0.857
	(0.381)	(0.593)

Note. This table reports the attrition rate in the parents' survey responses. In total, we received 1,882 valid responses, of which 1,852 were completed by students' parents, and the other 30 were completed by their other guardians (mostly grandparents). Panel A reports the attrition rate and the difference between the treatment and control groups. Panel B reports the degree of students' characteristics for predicting parents' nonresponses (Column 1) and tests whether the differences are correlated with the treatment assignment (Column 2). Classroom-level clustered standard errors are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

Table C5: Determinants of Take-up

	(1) OLS	(2) OLS
Family income		
lower middle	0.019	0.014
	(0.040)	(0.038)
upper middle	-0.011	-0.052
	(0.043)	(0.036)
high income	-0.045	-0.069
	(0.061)	(0.060)
Child age	0.242***	-0.028
	(0.055)	(0.062)
Victims	-0.066**	-0.056*
	(0.033)	(0.032)
Self-worth	-0.025*	-0.026*
	(0.015)	(0.015)
Depression	-0.111*	-0.083
	(0.058)	(0.062)
Stress	-0.019*	-0.016
	(0.010)	(0.011)
Parent's demographics	Yes	Yes
Child's demographics	Yes	Yes
Strata FE	No	Yes
N	1,004	1,004

Note. This table reports the results of the selection-into-participation pattern using the take-up dummy as outcomes with child and parent characteristics as regressors. The table only reports a few statistically significant determinants while leaving others in "child characteristics" and "parent demographics." Parent and child demographics are variables listed in Appendix Table C4. Classroom-level clustered standard errors are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

Table C6: Attrition and Robustness of Main ITT Estimates

	(1)	(2)	(3)	(4)	(5)
				Lee bounds	
	ITT	IPW	ITT (no FE)	Lower	Upper
Panel A. Mental Health					
GHQ score (Likert)	0.779***	0.847***	0.666**	0.659**	0.684**
	(0.282)	(0.274)	(0.296)	(0.322)	(0.338)
GHQ score (Standardized)	0.166***	0.181***	0.142**	0.140**	0.142*
	(0.060)	(0.058)	(0.063)	(0.069)	(0.073)
Feel very happy last week	0.064**	0.083**	0.050*	0.049*	0.052*
	(0.032)	(0.034)	(0.028)	(0.029)	(0.029)
Strata FE	Yes	Yes	No	No	No
N	1,852	1,852	1,852	2,246	2,246
Panel B. Three dimensions					
Social dysfunction	0.151**	0.157***	0.125*	0.125*	0.138*
	(0.059)	(0.057)	(0.064)	(0.068)	(0.070)
Anxiety	0.153**	0.178***	0.110*	0.106	0.122
	(0.062)	(0.062)	(0.064)	(0.070)	(0.074)
Loss of confidence	0.115**	0.126**	0.102*	0.102**	0.109**
	(0.051)	(0.049)	(0.053)	(0.065)	(0.067)
Strata FE	Yes	Yes	No	No	No
N	1,852	1,852	1,852	2,246	2,246

Note. This table reports the robustness tests to account for attrition in the parents' survey responses. Column 1 reports the ITT estimates, which are related to the results in Column 2 of Table 2. These estimates are calculated using equation (1), controlling for parents' demographic variables. Column 2 estimates the same equation using the inverse probability weighting method to account for attrition probability (Wooldridge, 2010). Column 3 reports the ITT estimates without the strata fixed effects. Columns 4-5 present the lower and upper bounds following Lee (2009)'s treatment effect bounds. Trimming proportion is 0.0011. In the Lee (2009) procedure, we standardize the GHQ and three dimensions without centering the variables and carry out the estimation without covariates. We perform 500 clustered bootstrap replications for standard errors. Standard errors reported in parenthesis are clustered at the classroom level. (p<0.10, ** p<0.05, *** p<0.01).*

Table C7: Robustness Checks: Alternative Scoring Methods and Cut-off Values

	(1) Control	(2) ITT	(3)	(4) TOT
Panel A. GHQ scoring methods				
GHQ (Likert)	0.000	0.169***		0.403***
	(1.000)	(0.062)		(0.151)
GHQ (0011)	0.000	0.087*		0.208*
	(1.000)	(0.048)		(0.113)
C-GHQ	0.000	0.086*		0.206*
	(1.000)	(0.050)		(0.118)
N	848	1,852		1,852
Panel B: Mental illness using various cut-off values				
	Control	ITT	% change	TOT
GHQ $\leq$ 21	0.091	-0.028**	-30.8 %	-0.066**
	(0.287)	(0.012)		(0.029)
GHQ $\leq$ 22	0.131	-0.031*	-23.7 %	-0.073*
	(0.337)	(0.016)		(0.040)
GHQ $\leq$ 23	0.175	-0.028	-15.1 %	-0.066
	(0.380)	(0.018)		(0.043)
GHQ $\leq$ 24	0.228	-0.029	-12.7 %	-0.070
	(0.420)	(0.020)		(0.048)
GHQ $\leq$ 25	0.318	-0.054**	-17.0 %	-0.130**
	(0.466)	(0.026)		(0.064)
GHQ $\leq$ 26	0.387	-0.059**	-15.2 %	-0.142**
	(0.487)	(0.025)		(0.064)
GHQ $\leq$ 27	0.448	-0.059**	-13.2 %	-0.141**
	(0.498)	(0.028)		(0.068)
GHQ $\leq$ 28	0.538	-0.077***	-14.3 %	-0.184***
	(0.499)	(0.029)		(0.071)
N	848	1,852		1,852

Note. This table shows the results of robustness checks of program effects on parental mental health, measured by GHQ-12. We try alternative scoring methods (Panel A) and test the results based on different cut-off values for the indicator of being depressed (Panel B). We use alternative scoring methods: To make them comparable, we standardized three total scores. ***GHQ (Likert)*** *adds up the original score (0-1-2-3);* ***GHQ scoring*** *uses the coding "0-0-1-1" for all items; and* ***C-GHQ*** *uses "0-0-1-1" for positive items and "0-1-1-1" for negative items. In Panel B, we report the results for different cut-offs that diagnose* ***Depression*** *as a mental illness, the cut-offs range from 21 to 28, and the most preferred cut-off value among existing literature is 24 or 25. Column 1 reports the means and the standard deviations for the control group. Column 2 reports ITT estimates. Column 3 in Panel B calculates the % change in depression rate, which is equal to the ratio of the ITT estimates to the control mean. Column 4 reports the TOT estimates. We controlled for strata fixed effects to take into account of stratified randomization design. First-stage Kleibergen-Paap rk Wald F statistic = 108.624. Classroom-level clustered standard errors are presented in parentheses (* p<0.10, ** p<0.05, *** p<0.01).*

Table C8: Romano-Wolf Adjusted P-values

	(1) Effect size	(2) RW p-values
Low income	0.864	0.026
Middle income	0.759	0.388
Highincome	0.634	0.388
Parent age (young)	0.962	0.041
Parent age (old)	0.716	0.125
Mother	0.751	0.034
Father	1.120	0.140
Child age (young)	1.056	0.006
Child age (old)	0.508	0.388
Child male	0.818	0.068
Child female	0.757	0.074
Child study pressure (low)	0.773	0.068
Child study pressure (high)	0.770	0.158
Child CESD (low)	0.899	0.020
Child CESD (high)	0.544	0.312
Onlychild	0.503	0.388
Have siblings	0.907	0.017

Note. This table complements Figure C7 with the Romano-Wolf adjusted P-values for the heterogeneous analysis. In this analysis, we account for 16 hypothesis tests and compute the adjusted p-values. Column (1) reports the effect size (estimates shown in Figure C7) for each subgroup, and Column (2) reports the corresponding Romano-Wolf adjusted P-values.

Table C9: Program Impacts on Test Score

	(1) Test score	(2) Grade rank
Panel A. Average effect		
Control Mean	0.0235	-0.0284
	(0.987)	(1.003)
ITT	-0.009	0.011
	(0.015)	(0.016)
Panel B. Quantile treatment effect		
1st Decile	0.010	0.010
	(0.020)	(0.024)
3rd Decile	-0.017	0.008
	(0.016)	(0.015)
Median	-0.012	0.013
	(0.014)	(0.013)
7th Decile	-0.013	0.017
	(0.014)	(0.016)
9th Decile	-0.010	-0.012
	(0.017)	(0.019)

Note. This table shows the program's effects on students' test scores, measured by the test score of the final exam, shown in Column 1, and grade rank, shown in Column 2. Panel A reports the average effects by first reporting the means and standard deviations in the control group and the ITT effects. Panel B reports the estimation results of an unconditional quantile regression. Classroom-level clustered standard errors are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

Table C10: Effects on Parents: Detailed Time and Monetary Investment

	(1) Control mean	(2) ITT	(3) Permutation test	(4) WCB
Pane A: Daily parental time inputs				
Weekday				
Read with child weekday	0.756	0.153**	0.009	0.014
	(0.899)	(0.058)		
Help homework	0.648	0.145**	0.010	0.012
	(0.933)	(0.055)		
Play with child	0.779	0.061	0.215	0.219
	(0.861)	(0.046)		
Conversation and family education	1.534	0.146*	0.096	0.101
	(1.418)	(0.083)		
Weekend				
Read with child weekday	1.011	0.161***	0.004	0.011
	(1.013)	(0.059)		
Help homework	0.882	0.158**	0.038	0.046
	(1.176)	(0.073)		
Play with child	1.257	0.011	0.683	0.852
	(1.006)	(0.054)		
Conversation and family education	2.254	0.040	0.718	0.704
	(1.663)	(0.097)		
N	848	1,852		
Panel B: Monetary input and attitudes				
Invest more than 1/4 income	0.491	-0.019	0.493	0.496
	(0.500)	(0.027)		
Tutoring if friend did	0.268	0.017	0.476	0.432
	(0.443)	(0.020)		
Tutoring if best student did	0.423	-0.001	0.959	0.956
	(0.494)	(0.023)		
Tutoring if most students did	0.463	-0.006	0.808	0.797
	(0.499)	(0.021)		
% believe tutoring helps in score	49.467	-0.135	0.888	0.894
	(20.203)	(0.985)		
% believe tutoring helps in mental health	46.890	-0.649	0.523	0.514
	(21.884)	(0.936)		
N	848	1,852		

Note. This table shows ITT estimates of Equation (1) for parental time and monetary investments. Panel A reports ITT estimates for different categories of time investments over the past week; Panel B reports parents' attitudes toward cram schools/after-school tutoring. Parents are asked to choose whether they will send their kids to cram schools in three hypothetical settings: Scenario 1 – when their best friends' children went to cram schools, Scenario 2 – when the best students in the class went to cram school, Scenario 3 – when most of the students in the class went to cram schools. Lastly, we elicit the perceived value of cram schools by asking parents to score (scale of 1-100) whether the cram school is good for students' test scores and whether cram school is good for students' mental health for a hypothetical struggling student. Column 1 reports the means and the standard deviations for outcomes for parents in control groups. Column 2 reports ITT estimates and standard errors, while Columns 3 and 4 report the associated permutation test and WCB p-values. Classroom-level clustered standard errors are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

Table C11: Effects on Time Spent with Parents, Reported by Children

	(1) Control mean	(2) ITT	(3) Permutation test	(4) WCB
Eat	3.025	0.218	0.151	0.164
	(2.364)	(0.144)		
Talk	3.548	0.419**	0.028	0.044
	(2.751)	(0.197)		
Watch TV	1.079	0.088	0.497	0.531
	(1.731)	(0.118)		
Check homework	1.739	0.576**	0.013	0.012
	(2.501)	(0.217)		
Outdoor activities	1.454	0.363**	0.020	0.033
	(1.917)	(0.164)		
N	1,029	2,246		

Note. This table shows the robustness test results when we analyze time spent with parents reported by students. The variables measure the number of particular events with parents during a normal week in the intervened semester (range 0-8, with 8 meaning more than 7 times). Column 1 reports the means and the standard deviations for students in control groups. Column 2 reports ITT estimates and standard errors, while Columns 3 and 4 report the associated permutation test and WCB p-values. Classroom-level clustered standard errors are presented in parentheses (p<0.10, ** p<0.05, *** p<0.01).*

Table C12: Summary of Cost-effectiveness in Literature

Panel A. Cost-effectiveness of mental health intervention				
Author	Region	Intervention	Cost	Size effect
Fuhr et al. (2019)	India	Peer support	$133/person or above	4.5 pp recovery
Richards et al. (2020)	UK	Internet-delivered CBT (iCBT)	£ 95/person	0.18 SD
Delgadillo et al. (2022)	UK	Stratified care (pre-screening)	£248/person	0.49 SD - 0.56 SD
Huang et al. (2020)	UK, China, et al. (Meta-analysis)	Peer support	-	0.37 SD

Panel B. Estimated cost of treating depression			
Author	Region	Methods and Contents	Estimated cost of depression
Christensen et al. (2020)	48 countries including US, UK, Germany, and the vast majority were from high-income countries (Meta-analysis)	Meta analysis	Total annual societal cost for mood disorder: $7,048 per patient
Ding et al. (2022)	Urban China	(1) Data comprised 5% random sample of claims data from China's Urban Basic Medical Insurance; (2) Direct medical cost; (3) January 2013-December 2016	Total annual direct medical cost of depression: $555.5 per patient
Xu et al. (2016)	Shandong province in China	(1) Data from the electronic health records of two psychiatric hospitals that consisted of 25,289 outpatients (10%) and inpatients (90%) who were diagnosed with a mental disorder; (2) Cost items included direct medical costs, direct non-medical costs, and indirect costs; (3) 2005-2013	Total annual cost of mental disorders: $3,665.4 per patient
Greenberg et al. (2021)	US	(1) Data comprised administrative claims data and National Survey on Drug Use and Health (NSDUH) data; (2) Evaluated the incremental economic burden of adults with major depressive disorder (MDD); (3) 2010-2018	Total annual cost: $7,822.6 per patient

Note. This table reports the cost-effectiveness analysis of similar programs targeting mental health. Panel A reports the cost-effectiveness analyses for related studies; Panel B reports the estimated costs of mental health problems in the literature, which helps convert the benefits of preventing depression into a dollar figure. For comparison purposes, we converted all US dollars to the current value.