

NBER WORKING PAPER SERIES

REDUCING BULLYING:
EVIDENCE FROM A PARENTAL INVOLVEMENT PROGRAM ON EMPATHY EDUCATION

Flavio Cunha
Qinyou Hu
Yiming Xia
Naibao Zhao

Working Paper 30827
<http://www.nber.org/papers/w30827>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2023

We thank Michelle Belot, Lance Lochner, Shiko Maruyama, Isabelle Perrigne, Miguel Sarzosa, Petra Todd, Weilong Zhang, and participants of the HCEO-IESR Virtual Summer School in Socioeconomic Inequality at Jinan University, the 2022 ESA Special Meeting by JILAEE, EEA-ESEM Milano 2022, and the 2022 SOLE Annual Meeting. All remaining errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Flavio Cunha, Qinyou Hu, Yiming Xia, and Naibao Zhao. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Reducing Bullying: Evidence from a Parental Involvement Program on Empathy Education
Flavio Cunha, Qinyou Hu, Yiming Xia, and Naibao Zhao
NBER Working Paper No. 30827
January 2023
JEL No. I10,I20,J24,O10

ABSTRACT

According to UNESCO, one-third of the world's youths are victims of bullying, which deteriorates academic performance and mental health, and increases suicide ideation and the risk of committing suicide. This paper analyzes a four-month parent-directed intervention designed to foster empathy in middle schoolers in China. Our implementation and evaluation study enrolled 2,246 seventh and eighth graders and their parents, whom we assigned, at the classroom level, to the control or intervention condition randomly. We measured, before and after the intervention, parental investments, children's empathy, and self-reported bullying perpetration and victimization incidents. Our analyses show that the intervention increased investments and empathy and reduced bullying incidents. In addition, we measured costs and found that it costs \$12.50 for our intervention to reduce one bullying incident. Our study offers a scalable and low-cost strategy that can inform public policy on bullying prevention in other similar settings.

Flavio Cunha
Department of Economics
Rice University
P.O. Box 1892
Houston, TX 77251-1892
and NBER
flaviocunha@rice.edu

Yiming Xia
Southwestern University of Finance and Economics
55 Guanghuacun St.
Qingyang District
Chengdu, Sich 610074
China
xiaym@swufe.edu.cn

Qinyou Hu
Department of Economics
Rice University
P.O. Box 1892
Houston, TX 77251-1892
qinyou.hu@rice.edu

Naibao Zhao
Research Institute of Economics and Management
Southwestern University of Finance and Economics
Chengdu 610074
China
nzhao@swufe.edu.cn

A randomized controlled trials registry entry is available at AEARCTR-0007079
A online appendix is available at <http://www.nber.org/data-appendix/w30827>

1 Introduction

According to [Olweus \(1978\)](#), bullying is a repetitive aggressive behavior marked by an imbalance of power between the two sides involved. A recent report has estimated that one-third of the world's youth are bullying victims ([UNESCO, 2019](#)). Furthermore, mounting evidence indicates that bullying negatively affects adolescents' academic performance and mental health, increases suicide ideation, and, thus, the risk of committing suicide ([Sarzosá, 2021](#); [Sarzosá and Urzúa, 2021](#); [Hinduja and Patchin, 2010](#); [Molcho et al., 2009](#)). Unfortunately, the adverse consequences of bullying can persist into adulthood ([Copeland et al., 2013](#)). Although bullying has been a problem for generations and today's adolescents experience bullying in increasingly intrusive ways (e.g., cyberbullying), we still lack scalable solutions to prevent this problem, despite the rise in school safety as a top national priority in many countries.¹ Therefore, there is a need to develop and implement preventive, scalable, and low-cost bullying policies.

The consequences of bullying are well-known. However, the causes of and solutions to school bullying remain underexplored. The literature in developmental psychology conjectures that empathy, that is, the ability to detect others' emotions, prevents school bullying incidents ([Jolliffe and Farrington, 2006](#)). A large body of research studies the role of empathy in promoting prosocial and inhibiting antisocial behaviors ([Eisenberg and Miller, 1987](#)). However, to our knowledge, no study has investigated whether a scalable intervention can reduce bullying by fostering empathy development.

We designed a low-cost, highly scalable, and parent-directed intervention to fill the research gap and inform policy. The intervention aims to foster empathy in middle schoolers by educating and coaching their parents. *Why parents?* A growing literature shows that parents and the home environment play a central role in children's social-emotional development ([Cunha and Heckman, 2008](#); [Cunha et al., 2010](#); [Bono et al., 2016](#); [Del Boca et al., 2017](#); [Attanasio et al., 2020](#)). *Why middle schoolers?* [Cunha and Heckman \(2007\)](#) argued that different human capital components undergo sensitive development periods in distinct stages of the lifecycle. For example, while the returns to investments in language comprehension skills occur in the early stages of the lifecycle (e.g., [Kuhl et al., 2003](#)), the returns to fostering skills such as empathy, prosociality, and perspective-taking tend to be higher in early to middle adolescent periods (e.g., [Steinberg, 2015](#)). In addition, school bullying peaks in early to middle adolescence ([Fitzpatrick et al., 2007](#)).

The critical insight from the psychology literature that low levels of empathy correlate with bullying behaviors ([Cook et al., 2010](#); [Jolliffe and Farrington, 2006](#)) forms our intervention's premise. For this reason, our intervention curriculum combines an education component that teaches parents about empathy with a coaching element that encourages parents to incorporate empathy into their daily parenting practices. The four-month intervention consists of themes covered through eight biweekly parent-child reading activities and four empathy-oriented movies.

¹For example, see the U.S. Federal Commission on School Safety (2018). In China, see the Law on Protection of Minors, which that country enacted in 2021 (China Ministry of Education, 2021).

Parents and students access these materials via a mobile app that records their progress.² The app contains a check-in feature that automatically measures the intervention parents' take-up and registers leakage to control parents.

We implemented this intervention in two middle schools in Yongkang, a county in southern China. The study sample consists of seventh and eighth graders from these two schools, which have 48 classes and 2,246 students. We conducted extensive fieldwork before and after the intervention and assessed parents' investments and children's empathy development and mental health. In addition, we measured the prevalence of five types of bullying behaviors: threatening, physical bullying, rumor spreading, social isolation, and cyberbullying. Unfortunately, we find that bullying is widespread in this setting: about 38 percent of the students admitted that they have been bullied by others in the semester immediately before the intervention.

We randomly assigned each of the 48 classrooms to the control or treatment conditions to evaluate the intervention's impacts on parents' investments, children's empathy and mental health, and bullying behaviors. We find that the intervention has a multifaceted effect on students and their parents. First, parents spend more time in empathy-related activities. In addition, we show that the increase in parental time in empathy-building activities does not crowd out the time (or monetary) investments to form other skills. Finally, parents become more empathetic and more likely to adopt an authoritative (or democratic) parenting style. Research shows that this parenting style correlates with better outcomes for children, such as greater empathy and emotional regulation levels (Sroufe, 2005), higher degrees of independence and self-efficacy (Givertz and Segrin, 2014), more intrinsic motivation for learning and better academic performance (Mattanah, 2001), higher levels of social development and mental health (Aquilino and Supple, 2001; Miklikowska and Hurme, 2011), and fewer behavioral problems (Grolnick and Pomerantz, 2009).

Second, the intervention helps children develop empathy. Students in treated classes are more likely to be prosocial and empathetic toward others. We use repeated measures of empathy to construct an empathy index. We find that the treated students' score in this index is 10% of a standard deviation (hereafter, SD) above the control students'.

Third, the intervention effectively prevents bullying. For example, the intervention reduces victims and bullies by 4.4 and 5.3 percentage points, respectively. In addition, the intervention reduces the likelihood that students will witness bullying incidents by 6.1 percentage points or be bystanders by 5.2 percentage points. As we show below, the impacts are exceptional from the standpoint of a cost-benefit analysis, and they are robust to alternative specifications.³ However, our intervention does not reduce all forms of bullying. Indeed, in our study, the intervention had the most significant impact on reducing physical bullying.

Our paper is among the first to meticulously examine the role of parental involvement in ado-

²We designed and embedded a mobile app into the most prevalent social media platform in China, *WeChat*, so that participants can easily access all the materials.

³We detect a similar pattern in alternative specifications—we vary the sets of controls: (1) demographics, (2) social desirability scale, and (3) survey completion time, and we expand the model to allow for misreporting. In addition, we use parental self-reports of bullying to cross-check our findings.

lescents' empathy skill formation and school bullying prevention. It innovates by implementing a large-scale parent-directed intervention in a real-world setting. Thus, our impact evaluation provides evidence of the effectiveness of scalable anti-bullying interventions. Because it is low cost, our parenting program does not require directly targeting potential bullies. In addition, it builds empathy, which correlates strongly with other desired prosocial behaviors such as altruism and tolerance.

Specifically, our study contributes to the following four strands of the literature. First, our paper contributes to the literature on the prevention of bullying behavior. Most studies focus on the impacts of bullying on students' (mostly victims') outcomes, such as mental health, social exclusion, and educational outcomes (Brown and Taylor, 2008; Eriksen et al., 2014; Ponzo, 2013).⁴ However, there is much less research on documenting the determinants of bullying behaviors and how to use such information to design prevention programs.⁵ Our paper is among the first few experiments to analyze the potential causal pathways of bullying behaviors among teenagers. It complements the research in developmental psychology and other fields of social science that have found large correlations between bullying perpetration and various individual or contextual characteristics (Álvarez-García et al., 2015), including low empathy (Farrington and Baldry, 2010; Cook et al., 2010), a distant relationship between children and parents (Li et al., 2019), and online classes during the COVID-19 pandemic (Bacher-Hicks et al., 2022).⁶ Our paper is closely related to that of Sarzosa and Urzúa (2021), who built and estimated an empirical model of endogenous bullying with multiple outcomes and latent skills. However, their analysis only focused on those being bullied due to data limitations. In contrast, our paper provides richer evidence on various types of bullying behaviors among both bullies and victims.⁷

Regarding the literature on bullying prevention programs, tackling bullying through family education has been narrowly explored in the literature and the field. Compared to school-based programs, which are usually costly and whose effects usually fade as the programs are terminated, our intervention has the additional advantage of being low cost and more likely to have long-run effects, as it changes students' skills and peer relationships (Farrington et al., 2017; Castillo-Eito et al., 2020).⁸

⁴Brown and Taylor (2008) and Eriksen et al. (2012) found a negative correlation between bullying and educational, behavioral, and wage outcomes. In recent studies, researchers have attempted to estimate the causal impacts of being bullied on educational and later-life outcomes. Eriksen et al. (2014) employed data from Denmark and suggested that being bullied significantly decreases the academic outcomes of 9th graders in the short run. Likewise, Ponzo (2013) used data from Italy to show that bullying victimization in students in elementary and middle school results in them achieving lower grades.

⁵Xia (2019) suggested that the within-classroom inequality of family socioeconomic status contributes to being a bully.

⁶Bacher-Hicks et al. (2022) used real-time Google search data during the COVID-19 pandemic in the U.S. and found that there is a positive correlation between online schooling and bullying incidents, especially cyberbullying incidents.

⁷We measure 5 types of bullying behaviors, including both direct and indirect bullying. Our paper also complements the research on cyberbullying behaviors in Nikolaou (2017), which relied on the exogenous state-year variation in implementing anti-cyberbullying laws to show the strong impact of cyberbullying on youth suicidal behaviors.

⁸Farrington et al. (2017) and Castillo-Eito et al. (2020) provided a summary on bullying prevention programs and found that most of the results from randomized control trials show modest or no effects, while before and after comparisons show large short-run effects.

Second, our paper strongly connects to the parenting literature. The experimental evidence of returns on time investment in noncognitive abilities is scarce. Existing studies on parental involvement have tended to focus more on improving cognitive and noncognitive skills in early childhood (Bono et al., 2016; Del Boca et al., 2017; Attanasio et al., 2020).⁹ This paper complements the literature by providing evidence of the importance of parental involvement in children’s development in later years, i.e., adolescence.¹⁰

Third, there is a strand of the literature on how parents’ time with children positively impacts children’s cognitive outcomes. For instance, Fiorini and Keane (2014) analyzed time use diaries from a large sample of children in the Longitudinal Study of Australian Children and found that children’s time spent on educational activities with parents is a highly productive input for cognitive skills. Villena-Roldan and Ríos-Aguilar (2012) used various instruments for parents’ time with their children and found that maternal time with a child directly affects their children’s math scores. Our paper adds to the literature by showing that noncognitive skills can also be cultivated and that parents who are more involved in the upbringing of their children have children with more favorable attitudes and traits. In contrast to papers relying on observational data to explore the role of parental involvement in children’s social-emotional skills (Moroni et al., 2019; Zumbuehl et al., 2021), we can provide causal evidence from a randomized experiment in a developing country context.

Fourth, our research also builds on the literature on empathy skills, particularly prosocial behaviors.¹¹ Kosse et al. (2020) evaluated a mentoring program targeting elementary school students to study the formation of prosociality skills. The study identified prosocial attachment figures and intense social interactions as the two main drivers of child prosociality, measured as altruism, trust, and other-regarding behavior in everyday life. Our research complements their study by exploring the other potential channel, i.e., the role of parental involvement, in guiding children’s behaviors and focusing on the opposite of prosociality, i.e., bullying behaviors. Our intervention also adapts elements from Alan et al. (2021), who implemented an educational program in the context of a massive influx of refugee children in Turkey.¹² They found that the intervention increased social cohesion between local and refugee students and enhanced prosocial behaviors. Although both studies aim to improve children’s empathy skills, our research is distinct from theirs in terms of the targeted age group, the intervention design, and the delivery format.¹³

⁹One can refer to Cunha et al. (2021) for a more systematic review, and Attanasio et al. (2020) for the most recent analysis.

¹⁰Our article is also distinct from Barrera-Orsorio et al. (2020), who studied parental involvement programs across 430 public schools in Mexico and found no effects on students’ educational outcomes. By targeting older children, our paper emphasizes cultivating children’s noncognitive skills.

¹¹Zaki (2020) suggested that high levels of empathy improve prosocial skills. Boisjoly et al. (2006) showed that empathy also helps reduce racial discrimination.

¹²Boucher et al. (2021) also analyzed the causal effect of a mixed ethnic program in Turkey targeting 5-year-old children; they built and structurally estimated a friendship network formation model to uncover the underlying mechanisms and found that interethnic exposure and language speaking skills are the two main drivers affecting social cohesion.

¹³First, their program mainly targets elementary school children, while ours tackles school bullying behavior and low levels of empathy among middle-school students. Second, our education program is parent-directed and emphasizes the interaction between parents and children rather than peer group interactions within the classroom. Third, our

We designed our study to follow the four transparent SANS conditions in List (2020). First, we select seventh and eighth graders in two schools (one public and one private) in a county (small city). Students in counties account for half of the total students in China. Our sample reflects that students are under high-stress levels from education and receive little time from their parents. Second, we have almost no attrition rate from the student side.¹⁴ Third, our parental involvement treatment is quite natural to parents, as many of them often receive homework tasks from teachers. Reading and talking to children are also considered part of the daily routine. The experimental sessions took place in the family and school settings, and all the decisions were real. Finally, the intervention costs almost nothing, only requiring teachers' effort to notify parents. Participants did not receive any monetary incentives for participating in the intervention, which helped enlarge the possibility of a scalable intervention. A cost-effectiveness analysis shows that organizing this program costs as low as \$12.50 to reduce one bully, \$15 per victim, \$10 per bully-victim. The total costs, which account for the opportunity costs of voluntary parental time inputs, are about 20 times less than the traditional whole-school approach bullying prevention programs, such as the famous Olweus Bullying Prevention Program (OBPP) (Olweus and Limber, 2010). The low cost makes it easy to scale up, even though the effect may be moderate or minor in different settings from our sample.

The remainder of the paper is organized as follows. Section 2 introduces the background of adolescents in Chinese counties and the intervention. Section 3 presents the measures used, experimental design, sample distributions, and empirical strategy for the treatment effect analysis. Section 4 shows the results, as well as the heterogeneous effects. Section 5 presents the additional analysis. Section 6 discusses the treatment effect magnitude and cost-effectiveness. Section 7 concludes the paper.

2 Background and the Program

2.1 Adolescents in Chinese Counties (Small Towns)

Students in the early grades of middle school, aged between 13 and 15, are transitioning between early and middle adolescence. They are considered rebellious and emotionally unstable. They are also often involved in school bullying. While most of the existing interventions in China focus on the school level (Chan and Wong, 2015), parents play a nonnegligible role in helping students navigate adolescence. Parents are overwhelmed by the educational arms race and often neglect students' social-emotional development and the school environment (Schmitz, 2011).¹⁵ Therefore, we hypothesize that poor social-emotional development may contribute to school bullying

intervention develops other desirable personality traits among adolescents, such as self-esteem, that go beyond empathy skills and prosociality.

¹⁴The main attrition comes from parents' nonresponse, with around a 17.5% attrition rate that is reasonably balanced across the treatment and control groups.

¹⁵See also <https://edition.cnn.com/2014/09/09/opinion/china-education-opinion/index.html>.

incidents. Compared with major cities, the situation is even worse in small counties, which is the focus of our study.

The study sample includes seventh and eighth graders in one public school and one private school. The two participating schools are in Yongkang, a county belonging to Jinhua city, Zhejiang Province. Yongkang has accommodated nearly one million residents and reached 103,163 RMB (about 15,000 USD) GDP per capita in 2020 (National Bureau of Statistics of China). Households in Yongkang are relatively wealthy compared to those in other counties in China. However, when looking at the education system in general, the students there still receive a test-oriented education, as do those in most regions of China. More importantly, baseline survey data show that parental time investment is, on average, low: 39% of students reported that their parents spent zero hours checking their homework. In contrast, 41% of students claimed that their parents spent zero hours on outdoor activities on a typical weekend. Since parental time investment is a crucial input of children's future development, especially in noncognitive skills, public policy must encourage parents to get more involved in their children's education.

Our intervention targeting secondary schools in counties in a developing country has additional implications. First, secondary schools in counties educate about half of China's students.¹⁶ Still, no specific study has ever drawn on them.

Second, schools in counties play an essential role in reducing the rural-urban education division. In Table D1, we compare student composition and parents' characteristics in schools in small counties with their counterparts in cities and rural villages/towns. There are significant differences along multiple dimensions. Compared with those in counties and rural areas, students in cities have a more advantaged background in terms of parental support. They also have higher levels of cognitive and noncognitive skills, creating another layer of inequality.

In China, counties are ranked at the third level of the administrative hierarchy. They are suburban areas that are partly urban and partly rural. Their schools reflect this reality. For example, in our study sample, 54% of the students are rural students. Hence, programs targeting students in Chinese counties might help reduce rural-urban education inequality.

Third, schools in counties usually cannot retain talented teachers and have limited infrastructure as mega-cities to attract the most talented teachers (Guang et al., 2010; Liu et al., 2022). Parents in counties are more exam-oriented and focus heavily on preparing their children for all levels of entrance exams.¹⁷ Students suffer from dual pressures as they internalize the limited resource constraints and extremely high parental expectations. These pressures result in an increased likelihood of experiencing stress, higher levels of depression, and school bullying among students in schools located in small counties. Table D2 illustrates the school bullying situation at the baseline. Physical bullying and verbal bullying (including threatening and spreading rumors) are pervasive; approximately 50% of students consider themselves victims, and almost 20%

¹⁶See at: http://www.gov.cn/zhengce/2021-12/16/content_5661148.htm.

¹⁷"Education in China is becoming increasingly unfair to the poor." 2021. *The Economist*: May 26. <http://www.economist.com/node/15980839> (accessed 12/12/2022).

classify themselves as bullies.

2.2 The Intervention

Our intervention is low-cost, highly scalable, and parent-directed. Figure D1 illustrates the theoretical framework of the intervention. The intervention content helps parents learn empathy and positive parenting skills. The intervention encourages parents to interact with the children to improve their mental health, foster noncognitive skill development, and correct misbehavior. Indeed, according to Simulation Theory from psychology (Preston and De Waal, 2002; De Vignemont and Singer, 2006), adolescents can develop their social-emotional skills, particularly empathy, by interacting with their parents. In addition, students can cultivate these skills by watching and discussing the vivid examples in the tasks.

Based on this framework, we design the intervention to cover four monthly themes, as illustrated in Figure D2. The themes are empathy, perspective-taking, tolerance (i.e., recognizing the value of various personality types), and the role of social-emotional skills in maintaining relationships with others. Table D3 summarizes individual family involvement tasks.¹⁸ For each monthly theme, we asked parents to read two biweekly articles and watch a movie. The first biweekly article educated parents about the concept and its importance. In contrast, the second biweekly article coached parents on “*how-to*” implement procedures to educate their children.¹⁹

At the beginning of each treatment month, parents received the link to the movie delivered by the class teachers. We asked parents to access the film via the link and watch it with their children. All the movies are related to positive parenting or empathy.²⁰

We delivered the program through a platform with a unique check-in feature to avoid and trace potential spillovers. For convenience, we embedded the platform in *WeChat*.²¹ All the tasks and materials were uploaded and delivered on the platform, which accurately recorded parents’ enrollment in the intervention with the check-in feature. Parents were asked to register on the platform using their children’s student IDs before they could start carrying out the tasks. Successfully registering on the platform enabled parents to access the biweekly articles and monthly movies during the intervention. Parents and students were asked to read the intervention’s articles and watch the assigned movies online together. Once a task was finished, students and parents were encouraged to submit a short reflection essay via the platform with no word limit.

¹⁸The content of reading relies on two books by American psychologists as references: “*The Power of Empathy: A Practical Guide to Creating Intimacy, Self-understanding, and Lasting Love in Your Life*” by A. Ciaramicoli and K. Ketcham and “*The Stress Solution: Using Empathy and Cognitive Behavioral Therapy to Reduce Anxiety and Develop Resilience*” by A. Ciaramicoli.

¹⁹For 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 into their parenting routines.

²⁰For instance, in the first month, we asked parents and children to watch the movie “Looking up.” See more details at [https://en.wikipedia.org/wiki/Looking_Up_\(film\)](https://en.wikipedia.org/wiki/Looking_Up_(film)).

²¹*WeChat*, like Facebook or WhatsApp in the United States, is the most popular social networking platform in China. Class teachers use *WeChat* groups to communicate with parents in all the classes of the study sample.

3 Methods

3.1 Recruitment

We recruited students through principals and class teachers. Participating classes included seventh and eighth graders in the selected public school and only eighth graders in the chosen private school. The number of eligible classes was 48, with more than 2,200 students. The average class size was 45 in the public school and 51 in the private school. We prepared the consent form with a brief program introduction. School principals informed the class teachers, who recruited students and parents on our behalf. Teachers distributed the consent forms and explained the program to students and their parents in a parent meeting in early January 2021. As a part of the recruitment process, teachers introduced the general aspects of the intervention (i.e., the reading and movie-watching activities), the importance of children’s empathy development, and informed the required time commitment to participate in the intervention. The information was also included on the consent form. Parents volunteered to participate in the study and intervention. Finally, 2,246 students and parents agreed to participate in the experiment.²²

3.2 Experimental Design

Figure 1 shows the timeline of the experiment starting in January 2021. We collected the baseline survey in mid-January after the parent meeting. Class teachers gathered students in computer lab rooms to complete the survey. In February, we randomly distributed half of the 48 classes into the treatment group and the other half into the control group. Therefore, each class is a cluster in this study. We used a stratified cluster randomization design. The study sample has four strata: seventh and eighth graders in the public school and “top classes” and standard classes (eighth graders) in the private school. Within each stratum, we randomly assigned classes to the treatment or control groups.²³ At the beginning of the spring semester, we delivered the notification messages about the tasks for the treatment group to the two principals, and they gathered the teachers in the treatment classes for a meeting. The teachers in the treatment classes learned about the intervention. They agreed to deliver the notification messages to parents through *WeChat* messages during the intervention period. The teachers expected no extra workload except for forwarding the reminder messages of the tasks prepared by our research team to the *WeChat* group.

In early March, parents in the 26 treatment classes received a message inviting them to register on the platform and start the first month of tasks. The teacher delivered each task message biweekly in the following four months via the treatment classes’ *WeChat* group. We also reminded

²²As the tasks are conveniently embedded in resources they use for their daily activities, almost all potential participants gave their consent to participate in this program. However, parents understood that even if they agreed to participate, they did not necessarily need to complete all tasks. In practice, almost 30% of them did not enroll in the intervention. Nevertheless, nearly all students completed the two surveys, and more than 80% of parents completed the follow-up surveys.

²³In practice, 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. Therefore, in the end, we have 26 treatment classes and 22 control classes.

teachers not to deliver the messages to other parents. The data verified that there was no accidental message leakage, discussed in detail in Section 3.4. This communication occurred regularly after dinner on Fridays, normally at 7:30 pm. We targeted this specific day and time for two reasons. First, the teachers do not have to work on the weekend, so that schedule worked well for them. Second, we wanted to encourage parents to engage with their children. We hypothesized that such an engagement would likely occur on the weekend. The task message briefly introduced the month's topic and encouraged parents to participate in the program. In addition, we also asked the treatment classes' teachers to send out a reminder to parents to finish the intervention tasks at the beginning of the third month. Students and parents in the control classes received no information during the intervention. The exception was an invitation to the follow-up survey in late June. The follow-up survey was conducted right after the end of the intervention.

3.3 Measurement and Data

We collected measures on students' social-emotional skills and bullying behaviors in the baseline and follow-up surveys. To improve the quality of these surveys, students completed them using computer-assisted self-interviewing (CASI) with the help of class teachers in the computer lab on campus. We informed students that we would keep their answers confidential to encourage unbiased answers.

We used the enrollment records to trace parents' actual take-up of the parent-child activity tasks of the program, which are biweekly data exported from the platform. We used school administrative data on students' test scores to measure students' academic performance covering two exams: one before the intervention and the other right at the end of the intervention.

Our data collection protocol produced low attrition rates in the student follow-up survey. The attrition consists of a few mistakes when reporting student ID or name, which we excluded from the sample (less than 0.5%). However, our attrition rate for parental responses is more significant. Table D4 shows that the nonresponse rate by parents accounts for about 17.5% of the student sample. However, the rates are almost identical between the treatment (17.5%) and the control groups (17.6%). The number of student samples and matched parent-child pairs total were 2,246 and 1,852, respectively.

The primary outcome is school bullying behavior. The secondary outcomes include empathy, parental involvement, mental health and stress, and personality traits.

Bullying Behaviors We use multiple questions with specific examples to cover five bullying domains: (1) threatening/verbal abuse for verbal bullying, (2) hitting/kicking for physical bullying, (3) lying and spreading rumors for social bullying, (4) social isolation for another dimension of social bullying, and (5) abusive or hurtful texts online for cyberbullying. Below, we explain how we use these repeated measures to address misreporting in bullying behaviors.²⁴

²⁴We restrict the focus to within-class bullying incidents. We do so because it is the norm that students in Chinese middle schools tend to stay in the same class for consecutive years until graduation.

Comparing results across different domains gave us more confidence in handling the Hawthorne reporting effects. One may expect a systematic improvement if the Hawthorne effects exist. Additionally, we include spectators' questions in the student survey²⁵ and parents' knowledge about their children being victims of the five bullying incidents in the parent survey.²⁶ From the third-person point of view, we expect the Hawthorne reporting effect to be negligible.

Our empirical analysis combines the five bullying behavior questions into three categorical variables. A bully is a bullying perpetrator. A victim is the victim of a bully. Finally, a bully-victim is an individual who is simultaneously a bully and a victim. We note that these categories are mutually exclusive. In addition, we also constructed cumulative measures for bullies and victims separately by counting the total number of events that happened, as the indicator may mask the degree of severity among the different types of bullying events.

The use of self-reported bullying behaviors raises some concerns, including validity and recall bias. However, research in criminology has discussed these concerns and provided empirical arguments that justify using self-reports as a measure of criminal involvement (Hindelang et al., 1979). In addition, these concerns are much less severe about bullying because bullying is not punishable by law and not always unacceptable or morally condemned among youths. Hence, the issues raised in crime research may be milder in the study of bullying. Additionally, Junger-Tas and Marshall (1999) suggested that the self-reported method is more valid and reliable with young adolescents than adults.

Empathy Skill We use a self-reported instrument for students' empathy skill and follow Alan et al. (2021) to measure the two dimensions at the baseline: *perspective taking* and *empathetic concern*.²⁷ However, our sample students are, on average, about 69 months older than students in Alan et al. (2021). Therefore, in the follow-up survey, we added *prosociality* to construct a more reliable measure of empathy for adolescents. The new measure is closer to the modified *Interpersonal Reactivity Index* in the psychology literature (Davis, 1983). To overcome the potential multiple hypothesis testing issues, we follow Anderson (2008) to construct an inverse covariance weighting empathy index.

Parental Investment Time investment is measured by hours per day on activities such as reading, checking homework, playing, and engaging in general education with children on weekdays and weekends, respectively, over the past week. For monetary investment, we asked parents to report their monthly education-related expenditures as a proportion of their total income. Parents chose one alternative among five categories: 5% or less, 5-10%, 10-25%, 25-50%, and greater than 50% of total income.

Mental Health and Stress We use the Center for the Epidemiological Studies of Depression

²⁵“Have you witnessed school bullying in the last semester?” and “What would you do if you saw someone being bullied?”

²⁶For example, we asked, in the parents' survey, “Has your child ever mentioned being threatened at school?” Similarly, we did the same for the other four bullying incident types.

²⁷Kamas and Preston (2021) discussed different types of measures of empathy. It concluded that a self-reported survey is a valid measure of empathy.

Short Form, or CES-D-10, to measure mental health. The questionnaire has ten items in a Likert scale format (Yang et al., 2018).²⁸ The depression indicator is generated with a threshold value of 12. The *inverse CESD index* is constructed by subtracting 30 from the CESD score so that a higher score indicates a better mental health status. We use a seven-item questionnaire to measure happiness. The higher the score, the happier the respondent. In addition, we measure source-specific student stress: (i) studies at school, (ii) peer relationships, (iii) rank/test scores in the class, and (iv) family background. We use a seven-point Likert scale for each item, from the least stressed (1) to the most stressed (7). We then construct a *mental health index* consisting of these measurements so that a higher score indicates that students are happier and less stressed.

Other Outcomes We also collect rich information on parents' inputs, empathy skills, and parenting styles. Tables D5 and D6 summarize the detailed measures we developed. Appendix Section A presents the survey questions on these outcomes in detail.

3.4 Summary Statistics

Baseline Characteristics Table 1 shows the descriptive statistics and tests the baseline balance between treatment and control groups. Panel A reports the demographic characteristics of students, and Panel B shows the baseline outcome variables, such as school bullying behaviors, depression, and social-emotional skills. Column (1) reports the summary statistics of the baseline outcome variables, while Column (3) shows the differences in students' characteristics between the control and treatment groups.

As shown in Panel A, students are, on average, 14.5 years old, and 53% of them are male. On average, urban residents consist of 46% of the sample. Having siblings is common in this sample, with 70% of students reporting having at least one sibling. The average height and weight of the sample are 161.9 cm and 50.5 kilograms, respectively. These figures are remarkably similar to the statistics from the nationally representative China Education Panel Survey (CEPS) (height 161.6 cm and weight 49.9 kilograms).²⁹ Compared with the summary statistics shown in Table D1, our sample has 6 percent more urban residents and 10 percent fewer only-child families.

Students have interactions with peers in the classroom. Many indicated five good friends in the friendship network question; the mean number of good friends is 4.³⁰ In terms of school bullying behaviors in the semester before the intervention, 70.6% of students reported at least one victimization experience, while 37.9% reported experiencing at least one type of perpetuation. Among the five types of school bullying behaviors, rumor spreading and physical bullying were the two top concerns.³¹ Table D7 shows that male students and students with low empathy skills

²⁸The items are also employed in the China Family Panel Studies (CFPS) 2012 survey.

²⁹The China Education Panel Survey (CEPS) is a large-scale, nationally representative, longitudinal survey starting with two cohorts of seventh and 9th graders in the 2013-2014 academic year. For more details, one can refer to <http://ceps.ruc.edu.cn/English/Home.htm>.

³⁰The number likely underestimates true counts because students could nominate at most five friends.

³¹Table D2 reports the exact proportions for each behavior: i. threatening other students (34% and 13%); ii. spreading rumors (53% and 17%); iii. physical bullying (45% and 18%); iv. social isolation (18% and 8%); and v. cyberbullying

and poor mental health are more likely to become bullies.

The total empathy score ranges between 9 and 63, and students on average score 48. Most of them express being somewhat empathetic in the survey. However, slightly more than 12% of them have a score less than half or 36.³² In terms of other personalities, students, on average, have a positive self-image. The average scores of feeling self-satisfied, having self-worth, being self-confident, and having self-esteem are all above four on a seven-point Likert scale. In addition, they are somewhat above average in perseverance, with a mean of 4.73 on the 7-point Likert scale. The average score on the final exam was 73% (460 out of 620), which they took the semester before the intervention.

Even though students were quite positive, they typically reported feeling stressed at the baseline. The stress scores ranged from 3 to 21. Students, on average, scored 13.2, and 56% of them had a score above 12. However, the average CES-D score was slightly below 9. If we use ten as the threshold for depression, 30% of students were screened as depressed, while 65% of students reported feelings of happiness in the past week, with an average score of 5.

We also found that parents and students were not close enough at the baseline; the total number of parent-child interactions within a typical week was 10.6. We report detailed categories of parent-child interactions in Column 1 in Table D8. The weekly average number of interactions indicated that students only dined with their parents three times and discussed their school lives with their parents slightly less than four times per week. In addition, parents took students to play outdoor activities about one and a half times and checked their homework assignments about two times.

Balancing Checks In Column 3 of Table 1, we test for imbalances in individual characteristics by reporting the mean differences between the treatment and control groups. We include demographics (Panel A) and pretreatment outcomes (Panel B). We then report standard errors for item-wise regressions using the variables listed in Column 1 as the dependent variables on the treatment indicator. The standard errors are clustered at the unit of randomization, the classroom, and reported in Column 4. All 22 estimates of individual regressions are nonsignificant, and these estimates are in line with what one would expect under random classroom assignments.

Actual Take-up As participation was purely voluntary, we only relied on the biweekly task messages to nudge the parents. Our platform records the actual registration and take-up of the tasks, and we use them to calculate biweekly task completion rates. In total, about 71% of eligible parents (N=872) registered in the platform and enrolled in the program.

Figure D4 shows the task-completion rates during the intervention. Overall, the task-completion rates gradually decreased. The rates experienced a significant increase around the middle exam after the teachers forwarded the midterm reminder. It then plummeted to about 15% in the final

(25% and 13%). Moreover, 63% of students have an exclusive, small group of friends, as shown in Table 1.

³²There is no standard threshold value for empathy measures, and we choose the medium value of the corresponding total score. We also show the distribution of the total empathy scores for the treatment and control groups in Figure D3.

month, as it was close to the time of preparation for final exams.

It is a four-month intervention, and most parents were not consistent enough to complete the program. If parents and students completed at least half the movies or readings, then we considered them as having participated in the empathy education, and we defined the rest of the parents as having not participated in the empathy education. This classification allowed us to study the effect of empathy education on students’ bullying behaviors. We also summed the number of movies and readings to investigate the dose effects of taking up empathy education.

The record allows us to verify potential spillover effects. Due to technology limitations, we were unable to make the link exclusive. Treated parents may have forwarded the link to others, or control parents obtained access to the platform. We checked the enrollment rates and found that this spillover problem is negligible. The take-up rate is zero in the control classes.³³ Therefore, we are confident that randomization at the class level is robust to spillover, as there is almost no communication between students in different classes. Our experiment meets the noninterference assumption, as we did not detect spillover between classrooms. Appendix Section B.1 discusses potential spillovers within classrooms in more detail.

3.5 Estimation of Program Effects

First, we estimate the intent-to-treat effect (ITT) by comparing outcomes across classes invited to participate in the education program (treatment) and other (control) classes. We follow the empirical specification as follows:³⁴

$$Y_{ic1} = \alpha + \beta_1 T_c + \beta_2 Y_{ic0} + \tau_s + \epsilon_{ic}, \quad (1)$$

where Y_{ic1} is a vector of outcome variables for individual i in class c at time 2 (follow-up), T_c is the treatment group indicator for class c , which was assigned at time 1 (baseline), Y_{ic0} is the baseline measures of the vector of outcome variables for individual i in class c , and τ_s is a set of strata fixed effects. In particular, we analyze the program’s effects on school bullying behaviors. The richness of our data also allows us to explore program effects on detailed types of bullying behaviors, including threatening, spreading gossip, physical bullying, social bullying, and cyberbullying. We cluster the standard error at the class level for all regressions using the Liang-Zeger estimator. Given 48 clusters, marginally greater than the rule of thumb, we complement it with Cameron et al. (2008)’s wild cluster bootstrap (WCB) p-values using 9,999 resamplings. To utilize the experimental nature, we also use a permutation test (Young, 2019) with 2,000 stratified clustered resampling.³⁵

³³Only four students (0.3%) in the control groups enrolled in the program but did not start any reading or movie tasks. Thus, the actual take-up rate of the control groups was zero.

³⁴McKenzie (2012) showed that in experiments with a single baseline and one follow-up survey, power is maximized when an end-line outcome is regressed on the treatment measure conditional on the baseline value of the outcome.

³⁵The computation of permutation p-value uses the exact small sample distribution of the test statistic. The method accounts for the complexity of the randomization and secures robust inference in small sample sizes (Walker et al., 2022). To alleviate the concern of relatively small clusters (48), we compute the wild cluster bootstrap (WCB) p-values.

Second, individuals in treatment classes were encouraged to participate in the program intervention, but they did so to varying degrees, as illustrated in Section 3.4. Highly motivated students, for instance, may fare well regardless of the program, but they may have a higher probability than other students of engaging in the program. Hence, we also estimate treatment-on-treated (TOT) models that account for individual students’ participation. We consider whether students and their parents actually “took up” the intervention, which is defined as having engaged in at least half of the reading or movie tasks. We estimate the TOT models using a two-stage least-squares (2SLS) design wherein the endogenous variable (the “take up”) is instrumented by the treatment assignment indicator with the actual take-up data exported from our platform. To fully use the actual take-up records, we also conduct a TOT-dosage analysis using the number of readings and movie activities for which participants registered. Following [Sylvia et al. \(2021\)](#), we estimate the 2SLS using the number of completed tasks as the regressor and control function approach, which we explain in detail in the Appendix Section C.1.

To better understand the effects of the intervention, we further explore the heterogeneous effects. We follow the generalized random forest (GRF) method proposed by [Athey et al. \(2019\)](#) to capture heterogeneity in a more disciplined way.³⁶ The advantage of the GRF method is that it can be easily applied to deal with high-dimensional baseline characteristics and predicts an importance rank for each variable based on its contribution to the heterogeneous treatment effects.

Furthermore, GRF extends the random forest approach by introducing and estimating a weighting function instead of simply averaging across all trees of a forest and using these weights to solve local moment equations. Once the importance rank is obtained, we then use these predictions to guide and conduct the traditional heterogeneity analysis based on the selected baseline characteristics. We discuss method details in Appendix Section C.2.

4 Results

The program evaluation follows the theoretical framework shown in Figure D1. We start with the direct effect of the program on parents’ involvement and parenting skills, which further affects students’ empathy skills, positive traits, and mental health. Improved students’ social-emotional skills lead to a better classroom climate with fewer bullies, victims, and bystanders in the treated classes.

4.1 Impacts on Parental Involvement and Outcomes

According to a general production function framework as in [Cunha and Heckman \(2007\)](#), children’s skills evolve as a function of parental investments and skills. Therefore, we investigate

³⁶For a more technical explanation of the GRF algorithm, one can refer to [Athey et al. \(2019\)](#). For a less technical description and examples of the application of the GRF algorithm to policy impact evaluations, one can refer to [Davis and Heller \(2017\)](#). We apply the ‘grf’ package in R to implement the algorithm: <https://cran.r-project.org/web/packages/grf/grf.pdf>.

whether our intervention changes these inputs before examining its impact on students' skills.

First, we explore the effects on different categories of time investment, including reading and talking with children, checking kids' homework, playing with kids, and educating them during the week and on weekends. This type of parental investment is crucial for child development. Then, we complement the study with the potential effects of the monetary investment (Del Boca et al., 2014). In particular, we investigate whether time spent with children crowds out (or in) investments in the form of expenditures on goods or services for the children.

Panel A in Table 2 reports the treatment effects on parents' investments. The estimates show that parents in the treated group significantly increased their time investment. The impact is more significant during weekdays than the effects on time investment on weekends. There is no significant impact on investment in goods or services.

Next, we estimate the program's effects on disaggregated measures of parent-child activities on weekdays versus weekends. First, the program led to a significant increase in time investment in children on weekdays. In addition, we find a substantial increase in reading with children and checking children's homework, as reported in Panel A in Table D9. This finding is consistent with a recent report by Doepke and Zilibotti (2021), which shows an increasing global trend in intensive parenting, especially in involvement in education-oriented activities. Although not significant in other parent-child activities, all the coefficients have positive signs. Additionally, Panel B in Table D9 shows that the program does not alter investment in goods or services. Furthermore, we do not find the program changes after-school tutorials or beliefs about them. We discuss more details in Appendix Section B.2.

The increased time investment is also due to increased take-up of empathy-related reading and activities. Table D10 presents the estimated ITT effects on the engagement of empathy-related parent-child activities. There is ample evidence that the program significantly increased engagement in empathy-related activities, including watching empathy-oriented movies and reading short articles on empathy. We measured the engagement frequency (i.e., at least once in the semester versus monthly during the semester). We detected that the most significant increase was in more frequent engagement.

Second, we study the program's effects on parents' skills. In our framework, parent skills include empathy skills and parenting style.³⁷

Panel B in Table 2 reports the effects on parents' empathy and parenting skills. The intervention led to a significant increase (0.10 SD) in the parents' empathy score compared to the control group. Furthermore, treated parents are 3.9 percentage points more likely to adjust their parenting style toward being more democratic. In addition, our program leads to an increase in demand for empathy education, suggested by a 5-percentage-point rise in willingness to invest in empathy education.

³⁷Simulation theory in psychology argues that empathic response is somewhat automatic and can be stimulated by imagining another person's affective state (Preston and De Waal, 2002; De Vignemont and Singer, 2006), we expect that the interaction between parents and children with the specific aim of developing empathy and prosociality may also affect the formation of the empathy skills of both.

4.2 Impacts on Students' Skills

Our parent-directed intervention aims to increase parental time and build parenting skills to improve students' social-emotional skills, particularly empathy, perspective-taking, and tolerance. We expect these improvements will reduce school bullying for three reasons that are not necessarily mutually exclusive. First, the psychological costs of bullying are an increasing function of an individual's social-emotional skills. Second, an individual's bullying motivation decreases with their social-emotional skills and increases with their stress levels. Third, an individual's degree of prosociality reduces their propensity to be bystanders. Hence, we test the program's effects on students' empathy skills, positive traits, and mental health.

Panel A in Table 3 reports that the program increases our social-emotional development index by 0.9 SD. The most significant improvements come from prosociality (0.16 SD) and empathy (0.10 SD). The TOT estimates suggest that program takeup (completing at least half of the tasks) increases prosociality, empathy, and perspective-taking by 0.4, 0.24, and 0.22 SD, respectively. The results are consistent with those in Table D11, indicating that the program significantly improves students' prosociality measured in hypothetical scenario questions.

Panel B reports that the program significantly improves the positive traits index, which summarizes improvements in self-esteem and perseverance. The TOT estimate suggests that taking up the program and finishing more than half of the tasks improve students' positive traits index by 0.34 SD.

Finally, Panel C shows that the program improves students' mental health as measured by combining the CES-D scale, the happiness scale, and the report on student stress. We attribute the increase in social-emotional skills and the reduction in stress to the increase in parental involvement.

Our parent-directed intervention educates and coaches parents to form their children's social-emotional skills. This format ensures that program takeup does not crowd out students' course time. Therefore, it should not produce any short-term negative effects on students' academic outcomes. In addition, the improvement in noncognitive skills may dynamically complement students' academic skills. However, it takes time for the program to have detectable impacts on test scores if there is dynamic complementarity between social-emotional skills and academic outcomes.

4.3 Impacts on Bullying

We hypothesize that improving students' social-emotional skills will reduce bullying incidents, as these skills correlate negatively with bullying propensity. Table D7 shows that students' baseline empathy skills, positive traits, and mental health are negatively correlated with their likelihood of being bullies and victims of bullying in both baseline and follow-up.³⁸

³⁸Furthermore, boys and students with more friends and who are members of exclusive groups are more likely to get involved in school bullying, which suggests that many school bullying incidents are likely to be collective behaviors; further studies incorporating students' networks would be valuable.

Table 4 reports the program's effects on children's bullying behavior. Column 2 reports the ITT estimates and standard errors using (1). The program reduces school bullying by significantly decreasing bullies, victims, bully-victims, witnesses, and bystanders in the treated classes. In these classes, we detect a 5.3 percentage points decrease in students who classified themselves as bullies, a 4.4 percentage points decrease for victims, and a 6.5 percentage points decrease for bully-victims. In addition, treated students are less likely to witness bullying incidents (6.1%) or to be bystanders (5.2%). Columns 3 and 4 report the permutation and WCB p-values. The inferences based on exact p-values are almost identical to the clustered robust standard errors. Finally, our results are robust to various specifications, including those allowing for misclassification errors (false negative rate), as discussed in Section 5.

We construct the cumulative bullying measure by considering the types of bullying behaviors, which has a value between 0 and 5. The difference between an indicator and the cumulative measure suggests that bullies and victims are often involved with more than one type of bullying behavior. The control group's bullying perpetration and victimization incident averages are 0.81 and 1.65, respectively. ITT estimates on the accumulative measure show a more significant reduction in bullying incidents for both bullies and victims.

From the perspective of observers, students in the treatment classes reported being less likely to witness bullying incidents. The intervention encouraged students to help victims. In addition, it reduces the likelihood that students will be bystanders. This result is scientifically significant because bystander is an essential predictor of bullying behavior (Salmivalli, 2010; Polanin et al., 2012).

We go beyond ITT estimates and study the effects of program take-up on school bullying behaviors. We consider the parents who completed at least half of the movies or half of the readings as a take-up group. Parents who only complete one of the tasks are more likely to enroll in the program and not really invest much time in it. Using the take-up indicator, we estimate the effect of taking up the program on the outcomes using 2SLS in Column 5 of Table 4. 2SLS reports larger effects, as the ITT estimates are generally a lower bound of the impacts. We show the TOT-dose regression estimates in Column 6 of Table 4. The does-response effects are all positive, and we detect significant linear effects of additional tasks on reducing the likelihood of becoming bullies and bully-victims. Students are also less likely to witness bullying incidents and to be bystanders. We detect marginally significant effects on lowering victimization.

The program may generate spillover effects via interactions between students or between parents. We discuss this in the Appendix Section B.1 and find that the spillover effects of taking up the program are modest. There is a significant negative correlation between the class-level take-up rate and individual bullying behaviors. Still, the significance faded once we instrumented the class-level take-up rate with the random assignment indicator.

Detailed Types of Bullying Incidents Bullying incidents can be categorized into two

groups: direct and indirect bullying.³⁹ Improving empathy may help explain the reduction in direct bullying because we can more easily identify the targets and the ways bullies operate (e.g., physical or verbal bullying). In contrast, the same channel may have a lower impact on indirect bullying (e.g., social isolation, rumor spreading, cyberbullying) because it is more challenging to witness and evoke peers' empathy.

To look beyond the aggregated bullying behaviors, we study the program's impacts on bullying behaviors by type. Panel A of Figure 2 presents the levels of the five types of bullying for treatment and control groups at follow-up using histograms. Almost 20% of the students admitted to spreading rumors or physically bullying others. The least frequent perpetration behavior is cyberbullying (11%). In addition, students reported a higher rate of being victims than bullies. For example, over 20% of the students were victims of cyberbullying and social isolation, and 50% of the students experienced being the target of rumors. The fraction of bully-victim is the lowest, as they are defined as simultaneously bullies and victims.

Comparing the levels between the treatment and control groups, treatment classes show lower levels in all five types of bullying behaviors. The difference between the levels of the types of bullying behaviors is either significant or marginally significant. We analyze these differences by estimating (1) for different types of bullying. Panel B of Figure 2 presents the ITT estimates (see exact numbers in Table D13). Panels A, B, and C in Table D13 display the effects on being a bully, a victim, and a bully-victim, respectively. The coefficients are all negative, regardless of the type of bullying behavior.

Our results suggest that reducing physical bullying and spreading rumors are the main drivers of the decrease in aggregated school bullying. Physical bullying is one of the most common types of bullying. Therefore, any reduction in its prevalence rate means that students experience a safer environment. Our parent-directed intervention reduced this type of bullying strongly. Indeed, the reductions in being a bully, a victim, and a bully-victim are 3.8, 5.5, and 3.5 percentage points, respectively.⁴⁰

Our intervention also reduced rumor spreading. For example, the intervention reduced the likelihood that an individual reported being a bully (3.8 percentage points) or a bully-victim (3.6 percentage points) of rumor spreading. Unfortunately, however, we find no differences in the propensity of being a victim of this type of bullying.

However, our intervention was arguably less successful in reducing other forms of bullying, such as cyberbullying, threatening, or social isolation.

³⁹According to van der Wal et al. (2003), direct bullying is a type of behavior that hurts, harms, or humiliates and is overt, obvious, and apparent to anyone witnessing it, whereas indirect bullying is not always immediately acknowledged as bullying. The different definitions often lead to different solutions and causes of these two types of bullying.

⁴⁰Table D14 shows our analysis of the intensive margin of bullying behaviors. The reduction in physical bullying is salient in terms of both perpetration and victimization.

4.4 Heterogeneous Treatment Effects on Bullying

Baseline Bully Status To better understand the “transition” of the bully status induced by the intervention, we investigate the program impact on bullying by students’ bully or victim status at baseline.⁴¹

Figure D6 shows the program impacts on the bullies (Panel A) and victims (Panel B) by four bully categories: being a bully, being a nonbully, being a victim, and being a nonvictim at baseline. From Panel A, we can see that our program significantly reduces bullying among nonbullies or nonvictims at baseline (although the difference is not significant). The program is more effective in lowering victimization among groups of students who were nonbullies and victims at baseline. Our program successfully helped the baseline victims get out of victimization. However, we find that the program effects on victimization are minor for baseline bullies. In contrast, the program effects on perpetration are smaller for baseline victims, which suggests that bullying perpetration and victimization are highly entangled.

Heterogeneous Treatment Effects Informed by GRF We can investigate the heterogeneity in treatment effects more systematically because we have a rich baseline dataset. For example, we measure maternal involvement, preprogram parental involvement, empathy skills, personality traits, ranks of test scores, associated pressures, and other social-emotional characteristics. Therefore, we apply the GRF algorithm to limit heterogeneity tests while minimizing the probability that important sources of heterogeneity are neglected. The GRF algorithm selects four baseline characteristics with the highest importance rank: empathy skill, age, parental involvement, and pressure score.⁴² It is well-established that male students are more likely to be involved in bullying incidents. Thus, we also test the impact of heterogeneity by gender.

Table D15 shows the heterogeneous effects on the indicator of being a bully (Panel A) and a victim (Panel B). Our analysis delivers policy-relevant insights. Our intervention produced a more significant reduction in the likelihood of being a bully for the children with low levels of parental investment at baseline. For example, the intervention reduced this probability by an extra 7.8 percentage points for the children in the bottom quartile of the distribution of parental involvement (on top of the gains experienced by the children in the top three quartiles of investment at baseline (Column 3). These findings are consistent with the literature on the importance of parental involvement in children’s development (Zumbuehl et al., 2021; Attanasio et al., 2020; Attanasio et al., 2020).

Similarly, we find that children with low baseline pressure scores benefited significantly more from the program than other children (Column 4). The average treatment effect on being a bully is 6.8 percentage points higher for children who had pressure scores in the lowest quartile at the start of the intervention compared with those in the top three quartiles. Studying the heterogeneous effects by pressure score contributes to understanding the unexpected consequences of China’s

⁴¹One advantage of our data is that we know the exact bully status, i.e., whether or not the student is a bully, at both baseline and follow-up.

⁴²For more detailed discussions, one can refer to the Appendix Section C.2.

extremely competitive education system on student outcomes (Guo and Qu, 2022; Jia and Li, 2021).

We continue to study the heterogeneous effects of being a bully-victim and bystander in Table D16. The results for being a bully-victim appear to deliver a similar pattern as those for being a bully. We find significantly higher treatment effects for those with lower parental involvement and lower pressure scores at the beginning of the intervention. We do not detect any differences along the other dimensions.

The results in Panel B show a heterogeneous effect on the willingness to help bullying victims. Compared to female students, we find that male students benefit more from the program, as they are less likely to be bystanders after the intervention. This finding suggests that male students may be less empathetic and, thus, experience higher treatment effects from the empathy education program.

We additionally explore the heterogeneous effects on students' empathy skills in Table D17. Although we hardly find any significant differences along the four dimensions, one thing to note here is that compared with female students, male students indeed have a lower level of empathy skills, and the heterogeneous effect by gender is marginally significant, as shown in Column 5.

In summary, the program reduces bullying among children with less parental involvement and less study pressure at baseline. Male students are less likely to be bystanders after the empathy education program. The impact is larger for children with low levels of family investments at baseline.

5 Robustness Analysis

5.1 Misreporting

The self-reported measures of bullying behaviors may suffer from misreporting because bullying is a socially non-desirable behavior. In this subsection, we explore the robustness of our analyses to allow for systematic misreporting.

Misclassification Error As suggested by Hausman et al. (1998) (HAS), we correct for the misclassification error in a binary choice model (i.e., when a response is reported in the wrong category). We describe the methodological approach in Appendix Section C.3. First, we additionally control for the baseline social desirability scale on top of the individual demographics to correct the potential bias. We show the results in Table D18. The odd columns show the treatment coefficients assuming that students answer truthfully, while the even columns display the HAS treatment coefficients assuming that there is a false negative reporting rate. Since we control for the baseline social desirability scale, the false negative rate is low. For bully and bully-victim indicators, the false negative rate is close to zero, while it is 0.092 for reporting as a victim.

Then, we follow De Paula et al. (2014) to report the treatment effects, assuming different levels of misreporting. The probit estimates are shown in Table D19. We can see that the patterns are pretty similar. All the specifications show negative treatment effects on bullying behaviors. We

find that misreporting leads to a downward bias of the treatment effect and that higher levels of misreporting lead to higher coefficient magnitudes. We also explore the misreporting of detailed bullying behaviors in Table D20. Again, we find a similar trend.

Hawthorne Effect Although there may be some concern about the Hawthorne effect, there are several reasons why this effect should be negligible. First, in this case, we should detect a significant decrease in all types of bullying. Furthermore, the spectators’ question about whether they “witnessed bullying incidents” also supports the absence of the Hawthorne effect. Second, the intervention materials never mentioned bullying, and the program was advertised as a parent-directed intervention to help children develop empathy. Thus, one should expect minimum experimenter demand effects on bullying outcomes. Third, teachers were not monitoring parent-child activities. They were not present during these activities and thus were only responsible for forwarding the text messages to the *WeChat* group of parents and had no access to content. Therefore, we expect almost no teacher effect on students’ reporting.

Finally, we conduct a cross-check by exploring the bullying incidents reported by parents. We asked parents whether their child had ever talked about being a victim of bullying. The ITT estimates shown in Table D21 confirm the program’s significant effect on bullying reduction; the program leads to a 3.1-percentage-point decrease for bullying victims. According to parents’ responses, rumor spreading and social isolation are the top two bullying behaviors that have the most significant reduction in terms of being a victim of bullying.

5.2 Alternative Specifications and More Evidence

We conduct several robustness checks in Table D22 to further support the main findings. In sum, the promising effects on school bullying reduction are robust to alternative specifications.

For ease of comparability, Column 1 reproduces the results from our main specification taken from Column 2 of Table 4. Column 2 of Table D22 reports the ITT estimates controlling for demographic characteristics, such as age, age squared, gender, and an indicator of being an only child. Column 3 further controls for survey completion length and its square. Column 4 reports estimates with additionally controlling for the social desirability bias index, following Dhar et al. (2022). Finally, Column 5 shows results from a pooled regression in which we view an individual type of bullying behavior as an observation and pool them together, thus generating a five times larger sample. We estimate the effects on the bullies, victims, and bully-victims with strata and type fixed effects and adjust standard errors clustered at the class and type levels. The results are similar across different specifications except for two cases. The ITT estimates for bullying victims become marginally significant once the social desirability scale is controlled for and when we conduct the pooled regression.

In Column 6, we also implement the entropy balancing (EB) method of Hainmueller (2012) to estimate TOT as a comparison.⁴³ EB is a multivariate reweighting method that makes the con-

⁴³This part of the analysis was conducted using the “*ebalance*” command in Stata following the instruction of Hainmueller and Xu (2013) and applying the default tolerance level of 0.015.

trol group data match the covariate moments of the treatment group.⁴⁴ Similarly to matching methods, EB can deal with selection on observables. However, EB has been shown to achieve a significantly higher level of covariate balancing than common propensity score approaches (Hainmueller, 2012).⁴⁵ From the results, we can see that EB gives more conservative estimates within the range of ITT and TOT.

We also investigate the treatment effects on detailed time use reported by the students as a form of robust evidence for parental time investment results in Table 2. The results are shown in Table D8. We can see that the program significantly leads students in the treated groups to be more likely to talk with and have their homework checked by their parents. Students in the treated groups also reported a significantly higher frequency of engaging in outdoor activities with their parents.

6 Magnitude and Cost-effectiveness Analysis

To position our program effect, we compare our program with existing bullying prevention programs. The majority of bullying prevention programs are school-based. The most well-known program is the Olweus Bullying Prevention Program (OBPP) (Olweus and Limber, 2010).⁴⁶ In a meta-analysis of 44 programs, Ttofi and Farrington (2011) showed that school-based bullying prevention programs, on average, reduce bullying prevalence by 17 - 23%.⁴⁷ In a recent and comprehensive review of 100 studies, Gaffney et al. (2019) showed that anti-bullying programs, in general, led to an approximate reduction of 19–20% and 15–16% for bullies and victims, respectively. Our program achieved a moderate effect in reduction of about 16% and 10% for bullies and victims, respectively. In terms of the reduction of bully-victims, the effect is substantial at about 22%.⁴⁸

⁴⁴The entropy balancing scheme assigns a scalar weight to observations in the control group such that the control group's distributions of all selected covariates match the treatment group's covariate distributions on the first and second moment. This strategy produces a sample in which the means and variances of all selected control variables are the same in the treatment and control groups. Of all the possible weighting schemes that fulfill these balancing requirements, entropy balancing chooses the one where all weights are nonnegative and deviate the least from uniform weights.

⁴⁵In Table D23, we further show summary statistics for the variables used for EB. As a comparison, we also report the standardized difference in means and compare the results with the traditional propensity score matching (PSM) method. The table highlights the differences between the two methods. After entropy balancing, the standardized difference in the means of all covariates is below 5% (Column 7), the criterion for successful matching proposed by Caliendo and Kopeinig (2008), and performs much better than PSM.

⁴⁶The OBPP was the first comprehensive school approach program (whole-school policy) which emphasized the involvement of school at all levels, family, and sometimes society (Richard et al., 2012; Beckman and Svensson, 2015). Parents in OBPP received leaflets or letters at home that provided them with information about bullying and the intervention program. Parents were also invited to information nights held at participating schools.

⁴⁷Many programs reviewed in Ttofi and Farrington (2011) delivered null effects, such as the Befriending Intervention Program (relied mainly on a peer support model), the Australian Friendly Schools program (whole school), the American Social Skills Group Intervention (curriculum based), the UK Youth Matters (curriculum based) and several others. Some programs' effectiveness depends on their implementation. The anti-bullying laws (ABL) showed an overall null effect, but strong ABLs are associated with an 8 to 12% reduction in bullying among US high schools (Sabia and Bass, 2017).

⁴⁸In the paper, we mainly report percentage points reduction, we obtain the percentage reduction by transferring the ratio of percentage point changes over the proportion of bullies/victims/bully-victims in the control group. For instance,

Regarding empathy education programs, the closest intervention is the one conducted and evaluated by [Alan et al. \(2021\)](#), who studied the impact of an education program on perspective-taking by training teachers and by delivering three lecture hours per week throughout one academic year targeting 80 Turkey elementary schools. They found a significant treatment effect of 1.21 fewer violent perpetration events and 0.75 fewer victimization events for a 10-consecutive-school-day period, although they detected null effects on student and teacher self-reported bullying behaviors. Their intervention led to a 0.27 SD gain in perspective-taking and no significant improvement in empathetic concern. [Kosse et al. \(2020\)](#) studied a mentoring program for the cultivation of prosociality among elementary school children in Bonn and Cologne, Germany. The program involved weekly one-to-one mentor-mentee interactions and lasted for one year (about 92 hours on average per year). They found that treated children experienced a 0.27 SD improvement in prosociality, and the effect persisted 2 years after the end of the intervention. As such, the effect size of our program on improving adolescents' empathy is modest, but it has additional beneficial effects on parents' skills and parent-child relationships, which help students develop more positive personality traits and mental health.

As for cost-effectiveness, the widely used school-based bullying prevention programs tend to be more costly compared to parental involvement programs. [Beckman and Svensson \(2015\)](#) showed that OBPP costs 131,250 Swedish kronor (\approx \$11,899) to reduce one victim spread. [Persson et al. \(2018\)](#) analyzed the cost-effectiveness of the Finnish KiVa bullying prevention program and found the cost is 131,321 SEK (\approx \$11,914) per gained quality adjusted life years (QALY)⁴⁹ or 7,879 SEK (\approx \$715) for each additional victim-free year gained. In Panel D of Table 5, we list cost-effectiveness results for several famous bullying prevention programs. Our program is much more cost-effective than the typical intervention. We expand the details of our program cost calculation in Panels A-C in Table 5. The organizational costs are only \$0.65 per student, which is negligible compared with those school-based programs. Accounting for the opportunity costs of voluntary parental time inputs, the costs per treated student may be as high as \$25, which is still a fraction of school-based programs. Our program cost lies in between [Alan et al. \(2021\)](#) and [Kosse et al. \(2020\)](#). In terms of cost per unit of effectiveness, our program costs about \$500 to reduce one bully, \$600 per victim, and \$400 per bully-victim. We should also note that our designed app can be easily adapted to other contexts and repeatedly used, thereby enhancing its cost-effectiveness. In addition, treating parental time input as opportunity costs may overestimate the costs as parents may value parental time differently, as suggested in [Guryan et al. \(2008\)](#).

to calculate the percent change of bullies, we get 16% from $0.053/0.328$.

⁴⁹QALY are calculated as the time spent in a state of health adjusted for the 'quality' or 'utility' of this time. QALYs reflect changes only in 'health'; they do not capture potential wider non-health benefits, such as improved educational attainment ([Bonell et al., 2015](#)).

7 Conclusions

To tackle the issues of increasing school bullying and low social-emotional skills among adolescents, we conducted an experiment of a directed parental involvement program in two large middle schools in a Chinese county, targeting 2,246 seventh and eighth graders and their parents. Motivated by the research evidence that indicates bullying behavior is usually associated with a lack of parental care and low levels of empathy, our intervention comprises biweekly activities that last for four months to develop a closer relationship between parents and children. We encourage parents in the treatment groups to watch monthly empathy-oriented movies and read biweekly articles on empathy and positive parenting with their children.

With the monthly take-up rate being approximately 40%, the program is still moderately effective in reducing school bullying incidents and improving students' empathy. Additionally, the treated students are more stress-resilient, self-satisfied, and less likely to be depressed than the untreated students.

The program also improves the treated parents' empathy levels, which they manage to incorporate into their parenting behaviors; parents in the treated groups are found to experience a 3.9 percentage points increase in adopting a democratic parenting style.

Finally, the modest take-up rate provides a direction for future improvement. The program can be easily scaled up for several reasons: (i) the intervention is simple and easily regulated by a smartphone application, which incurs little cost; (ii) the program is purely voluntary and is not likely to affect parent-teacher relationships; (iii) the intervention only lasts for a semester; and (iv) the increased time spent on the program successfully nudges parents to spend more time with their children without crowding out monetary investments. The incorporation of an encouragement mechanism into this type of directed parental involvement program may generate larger effects on students' personality development and reduction in school bullying.

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.
- Álvarez-García, D., T. García, and J. C. Núñez (2015). Predictors of school bullying perpetration in adolescence: A systematic review. *Aggression and Violent Behavior* 23, 126–136.
- 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.
- Aquilino, W. S. and A. J. Supple (2001). Democracy begins at home: Democratic parenting and adolescents' support for democratic values. *Journal of Family Issues* 22(3), 289–308.
- Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics* 47(2), 1148–1178.
- 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.
- Attanasio, O., C. Meghir, and E. Nix (2020). Human capital development and parental investment in india. *The Review of Economic Studies* 87(6), 2511–2541.
- Bacher-Hicks, A., J. Goodman, J. G. Green, and M. K. Holt (2022). The covid-19 pandemic disrupted both school bullying and cyberbullying. *American Economic Review: Insights* 4(3), 353–70.
- 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.
- Beckman, L. and M. Svensson (2015). The cost-effectiveness of the olweus bullying prevention program: Results from a modelling study. *Journal of Adolescence* 45, 127–137.
- Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles (2006). Empathy or antipathy? the impact of diversity. *American Economic Review* 96(5), 1890–1905.
- Bonell, C., A. Fletcher, N. Fitzgerald-Yau, D. Hale, E. Allen, D. Elbourne, R. Jones, L. Bond, M. Wiggins, A. Miners, et al. (2015). Initiating change locally in bullying and aggression through the school environment (inclusive): pilot randomised controlled trial. *Health Technology Assessment* 19(53), 1–109.
- 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.
- Boucher, V., S. Tumen, M. Vlassopoulos, J. Wahba, and Y. Zenou (2021). Ethnic mixing in early childhood: Evidence from a randomized field experiment and a structural model. *IZA Discussion Paper*.

- Brown, S. and K. Taylor (2008). Bullying, education and earnings: evidence from the national child development study. *Economics of Education Review* 27(4), 387–401.
- Caliendo, M. and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1), 31–72.
- 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.
- Castillo-Eito, L., C. J. Armitage, P. Norman, M. R. Day, O. C. Dogru, and R. Rowe (2020). How can adolescent aggression be reduced? a multi-level meta-analysis. *Clinical Psychology Review* 78, 101853.
- Chan, D. W. (1985). The chinese version of the general health questionnaire: does language make a difference? *Psychological Medicine* 15(1), 147–155.
- Chan, H. C. O. and D. S. Wong (2015). Traditional school bullying and cyberbullying in chinese societies: Prevalence and a review of the whole-school intervention approach. *Aggression and Violent Behavior* 23, 98–108.
- Cook, C. R., K. R. Williams, N. G. Guerra, T. E. Kim, and S. Sadek (2010). Predictors of bullying and victimization in childhood and adolescence: a meta-analytic investigation. *School Psychology Quarterly* 25(2), 65.
- Copeland, W. E., D. Wolke, A. Angold, and E. J. Costello (2013). Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence. *JAMA Psychiatry* 70(4), 419–426.
- Crowne, D. P. and D. Marlowe (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 24(4), 349.
- 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., E. Nielsen, and B. Williams (2021). The econometrics of early childhood human capital and investments. *Annual Review of Economics* 13, 487–513.
- 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.
- Davis, J. and S. B. Heller (2017). Using causal forests to predict treatment heterogeneity: An application to summer jobs. *American Economic Review* 107(5), 546–50.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology* 44(1), 113.

- De Paula, Á., G. Shapira, and P. E. Todd (2014). How beliefs about hiv status affect risky behaviors: Evidence from malawi. *Journal of Applied Econometrics* 29(6), 944–964.
- De Vignemont, F. and T. Singer (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences* 10(10), 435–441.
- Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *Review of Economic Studies* 81(1), 137–185.
- 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.
- Delavande, A. and H.-P. Kohler (2016). Hiv/aids-related expectations and risky sexual behaviour in malawi. *The Review of Economic Studies* 83(1), 118–164.
- Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping adolescents’ gender attitudes: Evidence from a school-based experiment in india. *American Economic Review* 112(3), 899–927.
- Doepke, M. and F. Zilibotti (2021). Do rising returns to education justify “helicopter” parenting? *IZA World of Labor* 487.
- Eisenberg, N. and P. A. Miller (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin* 101(1), 91.
- Eriksen, T. L., H. S. Nielsen, and M. Simonsen (2012). The effects of bullying in elementary school. *IZA Discussion Paper*.
- Eriksen, T. L. M., H. S. Nielsen, and M. Simonsen (2014). Bullying in elementary school. *Journal of Human Resources* 49(4), 839–871.
- Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics* 133(4), 1645–1692.
- Farrington, D. and A. Baldry (2010). Individual risk factors for school bullying. *Journal of Aggression, Conflict and Peace Research* 2(1), 4–16.
- Farrington, D. P., H. Gaffney, F. Lösel, and M. M. Ttofi (2017). Systematic reviews of the effectiveness of developmental prevention programs in reducing delinquency, aggression, and bullying. *Aggression and Violent Behavior* 33, 91–106.
- Fiorini, M. and M. P. Keane (2014). How the allocation of children’s time affects cognitive and noncognitive development. *Journal of Labor Economics* 32(4), 787–836.
- Fitzpatrick, K. M., A. J. Dulin, and B. F. Piko (2007). Not just pushing and shoving: School bullying among african american adolescents. *Journal of School Health* 77(1), 16–22.
- Gaffney, H., D. P. Farrington, and M. M. Ttofi (2019). Examining the effectiveness of school-bullying intervention programs globally: A meta-analysis. *International Journal of Bullying Prevention* 1(1), 14–31.
- Givertz, M. and C. Segrin (2014). The association between overinvolved parenting and young adults’ self-efficacy, psychological entitlement, and family communication. *Communication Research* 41(8), 1111–1136.

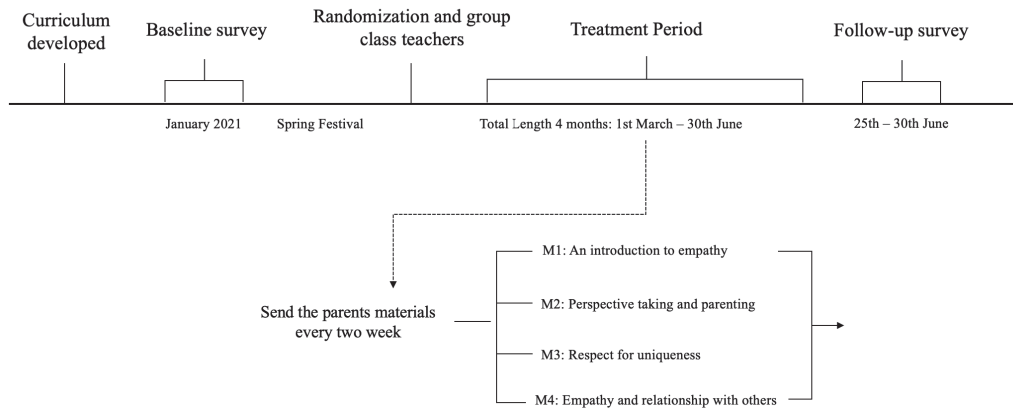
- Grolnick, W. S. and E. M. Pomerantz (2009). Issues and challenges in studying parental control: Toward a new conceptualization. *Child Development Perspectives* 3(3), 165–170.
- Guang, L., J. Adams, A. Gaetano, M. K. Whyte, B. Gustafsson, E. Hannum, H. Xiaojiang, F. Kong, L. Limei, L. Shi, et al. (2010). *One country, two societies: Rural-urban inequality in contemporary China*, Volume 16. Harvard University Press.
- Guo, J. and X. Qu (2022). Competition in household human capital investments: Strength, motivations and consequences. *Journal of Development Economics* 158, 102937.
- Guryan, J., E. Hurst, and M. Kearney (2008). Parental education and parental time with children. *Journal of Economic Perspectives* 22(3), 23–46.
- Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20(1), 25–46.
- Hainmueller, J. and Y. Xu (2013). Ebalance: A stata package for entropy balancing. *Journal of Statistical Software* 54(7).
- Hausman, J. A., J. Abrevaya, and F. M. Scott-Morton (1998). Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics* 87(2), 239–269.
- Hindelang, M. J., T. Hirschi, and J. G. Weis (1979). Correlates of delinquency: The illusion of discrepancy between self-report and official measures. *American Sociological Review*, 995–1014.
- Hinduja, S. and J. W. Patchin (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research* 14(3), 206–221.
- Jia, R. and H. Li (2021). Just above the exam cutoff score: Elite college admission and wages in china. *Journal of Public Economics* 196, 104371.
- Jolliffe, D. and D. P. Farrington (2006). Examining the relationship between low empathy and bullying. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression* 32(6), 540–550.
- Junger-Tas, J. and I. H. Marshall (1999). The self-report methodology in crime research. *Crime and Justice* 25, 291–367.
- Kamas, L. and A. Preston (2021). Empathy, gender, and prosocial behavior. *Journal of Behavioral and Experimental Economics* 92, 101654.
- Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch, and A. Falk (2020). The formation of prosociality: causal evidence on the role of social environment. *Journal of Political Economy* 128(2), 434–467.
- Kuhl, P. K., F.-M. Tsao, and H.-M. Liu (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences* 100(15), 9096–9101.
- Li, J., A. M. Sidibe, X. Shen, and T. Hesketh (2019). Incidence, risk factors and psychosomatic symptoms for traditional bullying and cyberbullying in chinese adolescents. *Children and Youth Services Review* 107, 104511.

- List, J. A. (2020). Non est disputandum de generalizability? a glimpse into the external validity trial. Working Paper 27535, National Bureau of Economic Research.
- Liu, A.-n., N. Liu, and A.-q. Wang (2022). Why can't rural schools retain young teachers? an analysis of the professional development of rural school teachers in china: Taking teachers in rural western china. *Social Sciences & Humanities Open* 5(1), 100238.
- Mattanah, J. F. (2001). Parental psychological autonomy and children's academic competence and behavioral adjustment in late childhood: More than just limit-setting and warmth. *Merrill-Palmer Quarterly* 47(3), 355–376.
- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. *Journal of Development Economics* 99(2), 210–221.
- Miklikowska, M. and H. Hurme (2011). Democracy begins at home: Democratic parenting and adolescents' support for democratic values. *European Journal of Developmental Psychology* 8(5), 541–557.
- Miller, P. H., S. D. Baxter, J. A. Royer, D. B. Hitchcock, A. F. Smith, K. L. Collins, C. H. Guinn, A. L. Smith, M. P. Puryear, K. K. Vaadi, et al. (2015). Children's social desirability: Effects of test assessment mode. *Personality and Individual Differences* 83, 85–90.
- Molcho, M., W. Craig, P. Due, W. Pickett, Y. Harel-Fisch, and M. Overpeck (2009). Cross-national time trends in bullying behaviour 1994–2006: findings from europe and north america. *International Journal of Public Health* 54(2), 225–234.
- Moroni, G., C. Nicoletti, and E. Tominey (2019). Child socio-emotional skills: The role of parental inputs. *IZA Discussion Paper*.
- Nikolaou, D. (2017). Does cyberbullying impact youth suicidal behaviors? *Journal of Health Economics* 56, 30–46.
- Olweus, D. (1978). *Aggression in the schools: Bullies and whipping boys*. Hemisphere.
- Olweus, D. and S. P. Limber (2010). Bullying in school: evaluation and dissemination of the olweus bullying prevention program. *American Journal of Orthopsychiatry* 80(1), 124.
- Persson, M., L. Wennberg, L. Beckman, C. Salmivalli, and M. Svensson (2018). The cost-effectiveness of the kiva antibullying program: results from a decision-analytic model. *Prevention Science* 19(6), 728–737.
- Polanin, J. R., D. L. Espelage, and T. D. Pigott (2012). A meta-analysis of school-based bullying prevention programs' effects on bystander intervention behavior. *School Psychology Review* 41(1), 47–65.
- Ponzo, M. (2013). Does bullying reduce educational achievement? an evaluation using matching estimators. *Journal of Policy Modeling* 35(6), 1057–1078.
- Preston, S. D. and F. B. De Waal (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences* 25(1), 1–20.
- Richard, J. F., B. H. Schneider, and P. Mallet (2012). Revisiting the whole-school approach to bullying: Really looking at the whole school. *School Psychology International* 33(3), 263–284.

- Sabia, J. J. and B. Bass (2017). Do anti-bullying laws work? new evidence on school safety and youth violence. *Journal of Population Economics* 30(2), 473–502.
- Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior* 15(2), 112–120.
- Sarzosa, M. (2021). Victimization and skill accumulation: The case of school bullying. *Journal of Human Resources*, 0819–10371R2.
- Sarzosa, M. and S. Urzúa (2021). Bullying among adolescents: The role of skills. *Quantitative Economics* 12(3), 945–980.
- Schmitz, R. (2011). The downside of exam-based education in china. *Marketplace*. Retrieved from <https://www.marketplace.org/2011/06/07/downside-exam-based-education-china/>.
- Sroufe, L. A. (2005). Attachment and development: A prospective, longitudinal study from birth to adulthood. *Attachment & Human Development* 7(4), 349–367.
- Steinberg, L. (2015). *Age of Opportunity*. Harper Paperbacks.
- Sylvia, S., N. Warrinnier, R. Luo, A. Yue, O. Attanasio, A. Medina, and S. Rozelle (2021). From Quantity to Quality: Delivering a Home-Based Parenting Intervention Through China’s Family Planning Cadres. *The Economic Journal* 131(635), 1365–1400.
- Ttofi, M. M. and D. P. Farrington (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology* 7(1), 27–56.
- UNESCO (2019). *Behind the numbers: Ending school violence and bullying*. UNICEF.
- van der Wal, M. F., C. A. M. de Wit, and R. A. Hirasing (2003, 06). Psychosocial Health Among Young Victims and Offenders of Direct and Indirect Bullying. *Pediatrics* 111(6), 1312–1317.
- Villena-Roldan, B. and C. Ríos-Aguilar (2012). Causal effects of maternal time-investment on children’s cognitive outcomes. *Center for Applied Economics, Working Paper* 285.
- Walker, S. P., S. M. Chang, A. S. Wright, R. Pinto, J. J. Heckman, and S. M. Grantham-McGregor (2022). Cognitive, psychosocial, and behaviour gains at age 31 years from the jamaica early childhood stimulation trial. *Journal of Child Psychology and Psychiatry* 63(6), 626–635.
- Xia, Y. (2019). What makes a bully? Available at SSRN 3683897.
- 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.
- 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.
- Zaki, J. (2020). Integrating empathy and interpersonal emotion regulation. *Annual Review of Psychology* 71, 517–540.

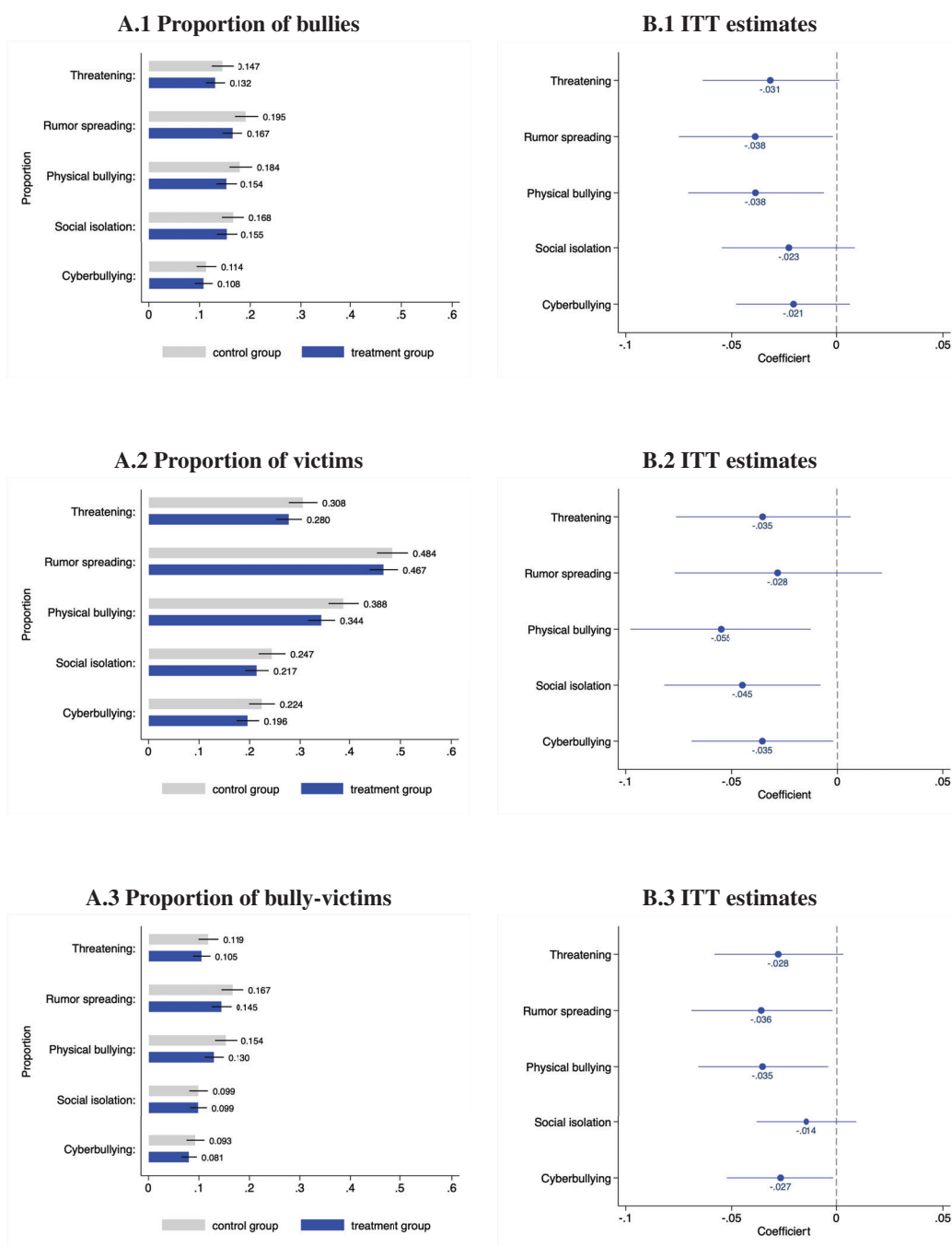
Zumbuehl, M., T. Dohmen, and G. Pfann (2021). Parental involvement and the intergenerational transmission of economic preferences, attitudes and personality traits. *The Economic Journal* 131(638), 2642–2670.

Figure 1: Timeline of the Experiment



Note. The top figure shows the timeline of the experiment. The intervention lasted for 4 months starting from March 1 until June 30. We list the monthly theme of the program during the treatment period.

Figure 2: Treatment Effects on Bullying Behaviors by Type



Note. (1) The figures on the left show the fractions of school bullying behaviors, including bullies in Panel A.1, victims in Panel A.2, and bully-victims in Panel A.3, for control and treatment groups at follow-up in detail. We document five detailed types of school bullying: i. threatening other students, ii. rumor spreading, iii. physical bullying (kick), iv. social isolation and v. cyberbullying. The numbers are calculated from the dummy variable of whether or not one was involved in certain types of bullying behavior in the past semester. (2) The figures on the right show the point estimates and 90% confidence intervals of the program's impacts on bullying behaviors in detail. The estimated effects are ITT estimates based on (1). The coefficient estimates are reported in Table D13. Confidence intervals are calculated based on robust standard errors clustered at the class level.

Table 1: Balance Test

VARIABLES	(1) Mean	(2) Std.dev	(3) Difference T-C	(4) S.E
Panel A. Demographics				
Age	14.463	0.578	-0.018	(0.136)
Male	0.530	0.499	0.008	(0.015)
Urban hukou	0.460	0.499	-0.028	(0.033)
Onlychild	0.298	0.458	0.007	(0.020)
Height in cm	161.885	7.799	0.182	(0.714)
Weight in half kilo	100.974	21.371	-1.171	(1.403)
Panel B. Outcomes at baseline				
Bully	0.379	0.485	-0.000	(0.025)
Victim	0.706	0.456	0.003	(0.023)
Number of friends	3.991	1.371	0.050	(0.130)
Member of exclusive group	0.628	0.483	0.015	(0.026)
Empathy score	48.120	9.636	0.935	(0.619)
Social desirability scale	6.193	2.017	0.128	(0.122)
Self-satisfied	4.461	1.884	0.112	(0.095)
Self-worth	4.776	1.752	0.108	(0.090)
Self-confident	5.000	1.692	0.064	(0.083)
Self-esteem	4.670	1.847	0.049	(0.089)
Perseverance	4.728	1.830	0.017	(0.086)
Total test score (620)	460.553	88.228	1.257	(8.529)
Stress score	13.229	4.149	-0.090	(0.260)
CESD 10-item	8.619	5.686	-0.547	(0.369)
Happiness score	5.067	1.735	0.134	(0.125)
Weekly interaction with parents	10.663	6.884	0.015	(0.769)

Note. (1) This table shows basic regressions attempting to verify the randomization of classroom assignments. Panel A reports demographic variables, while Panel B reports outcome variables at the baseline. (2) Columns 1 and 2 report the summary statistics for the whole sample. Column 3 reports the differences in means for each variable between treatment and control groups. Column 4 reports the standard errors for item-wise regressions using the variables labeled in the first column as the dependent variables. (3) Classroom-level clustered standard errors are presented in parentheses (* $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$).

Table 2: Effects on Parents' Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Control Mean	ITT	Permutation p-value	WCB	TOT	Dosage
Panel A: Time and monetary investment						
Time invest weekday	3.725 (3.288)	0.513** (0.204)	0.009	0.020	1.225** (0.483)	0.138** (0.055)
Time invest weekend	5.413 (3.649)	0.417* (0.230)	0.080	0.103	0.994* (0.544)	0.112* (0.062)
Money invest	3.395 (1.148)	0.012 (0.058)	0.838	0.842	0.029 (0.136)	0.003 (0.015)
N	848	1,852			1,852	1,852
Panel B: Parents' outcomes						
Very willing to invest in empathy education	0.585 (0.493)	0.050* (0.026)	0.073	0.073	0.119* (0.063)	0.013* (0.007)
Empathy index (short)	-0.036 (1.024)	0.103** (0.048)	0.050	0.040	0.245** (0.112)	0.028** (0.013)
Democratic parenting	0.789 (0.408)	0.039** (0.017)	0.033	0.033	0.094** (0.043)	0.011** (0.005)
N	848	1,852			1,852	1,852

Note. (1) This table shows the program impacts on parental investments and parents' outcomes using (1), 2SLS, and the control function approach. (2) Column 1 reports the means and the standard deviations for outcomes of control groups. Column 2 reports ITT estimates and standard errors, while Columns 3 and 4 report the associated permutation P-value after 2,000 stratified clustered resampling and wild cluster bootstrap P-value after 9,999 resampling. In Column 5, we report the TOT estimates using completing at least half of the reading or movie tasks as the "take-up" indicator and random assignment as an IV. In Column 6, we report the dosage-response estimates using the number of completed reading and movie tasks and a control function approach. (3) All regressions control for strata fixed effects. Classroom-level clustered standard errors are presented in parentheses (* $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$)

Table 3: Effects on Students' Outcomes

	(1) ITT	(2) Permutation p-value	(3) WCB	(4) TOT	(5) Dosage
Panel A: Empathy					
Empathy index	0.094** (0.044)	0.027	0.053	0.273** (0.135)	0.030* (0.015)
Empathetic concern	0.097 (0.060)	0.124	0.152	0.244* (0.141)	0.027* (0.016)
Perspective taking	0.086 (0.064)	0.180	0.225	0.219 (0.150)	0.024 (0.017)
Prosociality	0.160** (0.067)	0.012	0.019	0.402** (0.161)	0.045** (0.018)
Panel B: Positive traits					
Positive traits index	0.138*** (0.048)	0.003	0.007	0.339** (0.115)	0.038*** (0.013)
Panel C: Stress and mental health					
Mental health index	0.110** (0.051)	0.015	0.042	0.273** (0.120)	0.031** (0.014)
N	2,246			2,246	2,246

Note. (1) This table shows the estimated effects on students' empathy, positive traits, and mental health using (1), 2SLS, and the control function approach. The empathy index contains empathetic concern, perspective taking, and prosociality. The positive trait index contains self-satisfaction, self-worth, self-confidence, self-esteem, and perseverance. The mental health index contains the pressure score, happiness, and inverse CES-D. Refer to Table D11 for the impacts on the detailed components of each index. (2) Column 1 reports the ITT estimates and standard errors, while Columns 2 and 3 report the associated permutation P-value after 2,000 stratified clustered resampling and wild cluster bootstrap P-value after 9,999 resampling. In Column 4, we report the TOT estimates using those who completed at least half of the reading or movie tasks as the "take-up" indicator and a random assignment as an IV. In Column 5, we report the dosage-response estimates using the number of finished reading and movie tasks and a control function approach. (3) All regressions control for strata fixed effects. The default standard errors clustered at class-level are presented in parentheses (* $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$).

Table 4: Program Impacts on Bullying

	(1) Control Mean	(2) ITT	(3) Permutation p-value	(4) WCB	(5) TOT	(6) Dosage
Bullying involvements:						
<i>Indicator (0/1):</i>						
Bully	0.328 (0.470)	-0.053** (0.025)	0.038	0.044	-0.122** (0.059)	-0.013* (0.007)
Victim	0.619 (0.486)	-0.044 (0.029)	0.146	0.166	-0.107 (0.070)	-0.012 (0.008)
Bully-victim	0.295 (0.456)	-0.065** (0.025)	0.004	0.022	-0.138** (0.058)	-0.015** (0.007)
<i>Accumulative measure (0-5):</i>						
Bullying perpetration	0.808 (1.462)	-0.147* (0.083)	0.096	0.085	-0.367* (0.195)	-0.041* (0.022)
Bullying victimization	1.651 (1.726)	-0.194** (0.092)	0.044	0.054	-0.483** (0.217)	-0.054** (0.024)
Spectators:						
Witnessed bullying incidents	0.420 (0.494)	-0.061* (0.034)	0.092	0.094	-0.149* (0.079)	-0.016* (0.009)
Willing to help victims	0.767 (0.423)	0.052** (0.020)	0.014	0.009	0.129** (0.049)	0.014** (0.005)
N	1,029	2,246			2,246	2,246

Note. (1) This table shows the program impacts on school bullying using (1) 2SLS, and a control function approach. (2) Column 1 reports the means and the standard deviations for the corresponding outcome variables for students in control groups. Column 2 reports the ITT estimates and standard errors, while Columns 3 and 4 report the associated permutation P-value after 2,000 stratified clustered resampling and wild cluster bootstrap P-value after 9,999 resampling. In Column 5, we report the TOT estimates using completing at least half of the reading or movie tasks as the “take-up” indicator and random assignment as an IV. In Column 6, we report the dosage-response estimates using the number of finished reading and movie tasks and a control function approach. (3) All regressions control for strata fixed effects. Classroom-level clustered standard errors are presented in parentheses (* $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$).

Table 5: Cost-effective Analysis

Panel A: Benefits (program impact on bullying)	
Total benefit per class	2.4 bullies / 2 victims / 3 bully-victims
Total organizational costs	\$1,400
Organizational costs per class	\$30
Organizational costs per student	\$0.65
Organizational costs per unit of benefits	\$12.50 per bully / \$15 per victim spread / \$10 per bully-victim
Panel B: Organizational costs (include in-kind)	
Account for parental time input	\$1,170 per class
Costs per class (include parents' opportunity costs)	\$1,200 per class
Costs per unit of benefit	\$500 per bully / \$600 per victim spread / \$400 per bully-victim
Panel C: Opportunity costs	
Program Name	Costs
Olweus Bullying Prevention Program (OBPP) in Beckman and Svensson (2015)	131,250 Swedish kronor (\$11,899) per victim spared
Finnish bullying prevention program KiVa in Persson et al. (2018)	131,321 SEK (\$11,914) per quality-adjusted life-years (QALY) or 7879 SEK (\$715) for each additional victim-free year gained
Turkish refugee perspective taking program in Alan et al. (2021)	\$8.2 program cost per child
Germany prosociality mentoring program in Kosse et al. (2020)	€1,000 (\$822) per child per year
Panel D: Cost-effectiveness for other programs	

Note. This table reports the cost-effectiveness analysis of our program. Panel A reports the average benefit (reduction of bullying events) per class; Panel B reports the organizational costs; Panel C reports the costs taking into account potential opportunity costs of parental time input. Panel D reports the cost-effectiveness of several representative interventions for comparison purposes. One can refer to Gaffney et al. (2019) for a meta-analysis of the effectiveness of bullying prevention programs globally. All costs are converted into USD using the exchange rate of 7.2 RMB per USD. Total organizational costs are \$1,400 for 48 classes, treated classes did not occur extra costs as teachers only act as a messenger. In terms of opportunity costs, the hourly income is 1.5 × the local minimum wage (20 RMB). Total opportunity costs = hourly income × monthly input × number of compliers in each month = 30 × 4 × (20 × 3 + 10) = \$8,400 RMB (\$1,170).