Noncognitive Development and Peer Networks: Experimental Evidence on Chinese Rural Boarding School Students

Yue Li* Peiyi Jin†

March 3, 2026

Abstract

This paper evaluates the direct and spillover effects of a low-cost, scalable intervention- audio bedtime stories- on non-cognitive skill development among rural boarding school students in China. Using a randomized controlled trial (RCT), we analyze data from 63 schools, where boarding students were exposed to bedtime stories broadcast through dormitory speakers. This design isolates treated students from untreated day students in the same classroom, allowing for robust identification of peer effects in non-cognitive skill transmission. The results show significant positive effects on treated students, with larger gains for those with day-student friends and larger friendship networks. Spillover effects on untreated day students were primarily driven by those with boarding friends, highlighting the critical role of peer networks in the transmission of non-cognitive skills. Students treated earlier exhibited greater improvements in resilience. Additionally, we estimate the average peer effect of friends' non-cognitive outcomes on a student's own outcomes using a Linear-in-Means model, with the treatment acting as an exogenous shock to peer outcomes.

This study provides the first experimental evidence on the transfer of non-cognitive skills through peer networks in rural Chinese boarding schools. It is novel in terms of separating treated and non-treated students within the same classroom and see how skills transfer through peer networks, contrasting with previous studies that rely on random groupings or classroom-wide interventions. Our findings contribute by (1) offering a scalable intervention to enhance non-cognitive skills through peer networks, (2) demonstrating the amplification of treatment effects through social networks, and (3) advancing the study of peer effects using detailed network data. The program has reached over 4 million children in China, showcasing its widespread impact.

Keywords: Human Capital, Social Networks, Noncognitive Skills, Field Experiment

JEL classification: I21, I26, D85, C93, J24, O15

*University Collge London, yue.li.22@ucl.ac.uk.
†National University of Singapore, jin_peiyi@u.nus.edu.

We are grateful to Orazio Attanasio, Sumit Agarwal, Wenrui Huang, Haoming Liu, Jinfeng Liu, Jin Liang, Martin Mattson, Aureo de Paula, Imran Rasul, Yingquan Song, Chiyao Sun, Liyang Sun, Michela M. Tincani, Tiffany Tsai, Yaowu Wu, and participants in the The Society of Labor Economists 2024 Annual Meeting, 2024 Annual Meeting of the European Association of Young Economists, The 16th China Economic Association (Europe) and 35th China Economic Association (UK) Annual Conference, 2024 IFS-UCL-LSE/STICERD Development WiP Seminar, Junior Research Day 2024 Autumn London Meeting, 2024 NUS Applied Economics Student Workshop and 2023 UCL PhD Seminar for insightful comments. This paper is based on a field experiment and four rounds of data collection with researchers at the Chinese Academy of Social Sciences (CASS) Institute of Population and Labor Economics, China Institute for Educational Finance Research(CIEFR), and the Beijing Growing Home Foundation. Research funding from the National Natural Science Foundation of China's grant number 71973150 is gratefully acknowledged.

1 Introduction

Extensive empirical research has provided valuable insights into the effect of human capital on economic development and labor market outcomes. While cognitive skills have traditionally taken center stage in our understanding of intervention impacts, a growing body of evidence has highlighted the equally pivotal role played by non-cognitive skills, such as resilience and self-esteem (Kautz, Heckman, Diris, Ter Weel, and Borghans, 2014, Balart, Oosterveen, and Webbink, 2018, Heckman and Kautz, 2012). However, there remains a lack of compelling evidence regarding how these personality traits develop, particularly through peer interactions, and how interventions may influence not only treated individuals but also their peers.

This paper addresses these gaps by making three key contributions. First, it enriches the literature on interventions for non-cognitive skill development by evaluating both direct and spillover effects in the context of rural Chinese boarding schools. Through a randomized controlled trial, we investigate how non-cognitive skills such as resilience and self-esteem are influenced not only in treated students but also in their untreated peers via friendship networks. This study provides new empirical evidence on how peer outcomes, rather than just peer characteristics, affect the development of non-cognitive skills. While previous studies have predominantly focused on peer attributes like gender, race, or academic achievement (Hoxby, 2000, Figlio, 2007, Sacerdote, 2014), our research extends the understanding of how non-cognitive skills can be transferred within social networks. Our results indicate that students benefit more from being treated at an earlier stage in terms of resilience development, further emphasizing the importance of timely interventions during critical developmental periods.

Second, this paper highlights the importance of friendship networks as a channel for peer influence on non-cognitive outcomes. Notably, our findings show that the spillover effect on day students is primarily driven by those with boarder friends, and the treatment effects on boarding students were larger for those with more day-student friends and those with larger friendship networks. Interestingly, however, there does not appear to be a need to specifically target individuals with larger networks for intervention, as the effect is naturally reinforced through these networks. Treating everyone equally allows individuals with more extensive networks to receive additional benefits from the treatment, suggesting that broad, inclusive interventions can efficiently leverage existing peer relationships for maximum impact.

Third, this paper advances the methodological discussion on peer effects by incorporating detailed network data, providing a clearer picture of the dynamics at play in peer-to-peer skill transmission. Prior research has often been limited by the broad definitions of peer groups, typically using classmates or grade cohorts. Our study leverages rich network data from a large-scale intervention in rural Chinese schools, allowing us to more precisely identify friendship ties and determine which peers matter most in the transmission of non-cognitive skills. This detailed data helps refine our understanding of how peer networks can amplify the benefits of educational interventions, offering important

insights into how social connections contribute to the development of resilience and self-esteem.

To explore these mechanisms, we conducted a field experiment across 137 rural boarding schools in five counties in China. Students were randomly assigned to one of three treatment groups or two control groups. This paper focuses on the "audio bedtime story only" treatment group, where boarding students heard bedtime stories through speakers in their dormitories, and a control group. Previous research has demonstrated that bedtime storytelling facilitates children's socio-emotional and cognitive development, including self-discipline and planning abilities (Blake and Maiese, 2008, Hutton, Horowitz-Kraus, Mendelsohn, DeWitt, Holland, and Consortium, 2015). Our intervention, at a cost of approximately 1.15 pounds per student per month, represents a highly cost-effective method for fostering non-cognitive skills in rural settings.

We tracked students surveyed in 2015, and conducted follow-ups in 2016 and 2017 to evaluate the direct treatment effects and the spillover effects on untreated peers. We measured resilience, self-esteem, internalizing and externalizing behaviors, and depression using validated scales (Song, 2003, Rosenberg, 1965, Radloff, 1977). Our findings demonstrate that the intervention had a positive impact on personality traits, including resilience and self-esteem, but the effects take time to manifest. For example, resilience increased by 0.086 standard deviations after several months of treatment, with each intervention costing less than two pounds per month per student.

Additionally, this research contributes to policy discussions on improving student outcomes in low-income areas. In rural China, where many children are left behind by parents seeking work in urban areas, students often face emotional and behavioral challenges due to a lack of parental companionship (Hoeve, Stams, Van der Put, Dubas, Van der Laan, and Gerris, 2012). This study demonstrates that even low-cost interventions, such as providing audio bedtime stories, can have significant positive effects on personality development, both directly and through peer networks. These results offer actionable insights for educators and policymakers seeking to address the socio-emotional needs of rural boarding students.

In summary, this paper contributes to the literature in three key ways: (1) it enriches the understanding of non-cognitive skill interventions by evaluating both direct and spillover effects in a rural Chinese context; (2) it highlights the critical role of friendship networks in shaping peer influence on non-cognitive outcomes, demonstrating that treatment effects and spillover effects can naturally arise without needing to target students with larger networks; and (3) it advances the methodological discussion by incorporating detailed network data, providing a clearer understanding of how non-cognitive skills transfer through peer interactions. These contributions are essential for designing effective interventions that leverage social networks to improve educational and developmental outcomes for students in under-resourced environments.

2 Institutional Background

Since the number of students was declining, especially in rural areas of the central and western regions of China, many local governments executed a program known as "school consolidation and merging" between the years of 2001 and 2012. In rural areas, this program resulted in the consolidation of many teaching points. Due to the reduction in the number of elementary schools and their concentration in townships, pupils had to travel farther from their homes to attend school. As a result, younger elementary school students in the lower grades had to choose boarding accommodations. Till 2017, 10.66 million primary school students and 20.74 million junior secondary school students boarded, making up 10.6% of all primary school students and 46.7% of junior secondary school students. Particularly in western and central China, the numbers are larger in rural areas. According to China Census 2020, 36.4% of the children (0-17) don't live with both parents, In rural areas, 42.1% don't live with both parents.

Psychological literature suggests that a lack of parental companionship might cause emotional and behavioural disorders (Hoeve et al., 2012). Research on rural boarding students found that boarding has a significant negative impact on rural students' non-cognitive skills; boarding students are more likely to have anxiety, depression, and interpersonal problems; and their overall level of interpersonal trust is lower Song and Zhongchen (2018). Boarding school kids were also found to perform rather poorly in terms of social-emotional skills (Wang, Dong, and Mao, 2017). Parents leaving their children to go to work has a major detrimental impact on their social-emotional development, and the proportion of left-behind children among boarders is comparatively high. Most rural parents have a junior high school education or less. They frequently lack the knowledge and abilities to help their kids with their schoolwork, making it challenging to teach them non-cognitive skills like sociability. Parents of rural boarding students interact with their children less than those of urban boarding students, are less aware of their needs internally, and fail to notice changes in their internal feelings. Because of this, rural boarding students frequently struggle with deteriorating parent-child bonds, a lack of emotional support, and the growth of non-cognitive skills (Jingzhong and Lu, 2011).

More left-behind children have resulted from the growth of rural boarding schools, which has had a detrimental effect on kids' psychological and academic performance. It has had an impact on the popularity of high school education to some extent and raised the danger of junior high school dropout among pupils (Hou, Wu, Song, et al., 2018). Non-cognitive skills and cognitive skills have a close mutually supportive relationship, which is consistent with the idea that "skills produce skills"; specifically, the development of non-cognitive skills in one stage fosters the development of cognitive abilities in the following stage. Such reliance is also present in the rural primary and secondary institutions of western China (Glewwe, Huang, and Park, 2017). Rural boarding students' and left-behind children's development of non-cognitive skills could also have a direct impact on their academic performance in junior high and even high school, which is connected to their future and welfare as individuals and, more importantly, the need for talents in

social and economic development. Rural boarding schools are perfect study objectives if we want to target children lacking parent companionship.

The paper makes significant contributions across multiple domains of research. Firstly, it enriches the body of literature concerning interventions in child development. Previous research has demonstrated effective interventions aimed at enhancing students' cognitive abilities. For example, Gordanier, Hauk, and Sankaran, 2019 finds that additional support, including tutoring and skill training, substantially improves final exam scores, with particular advantage for students with below-average math placement scores. Recent research also underscores the efficacy of targeted interventions in molding socioemotional skills, as highlighted in studies by Baumert, Schmitt, Perugini, Johnson, Blum, Borkenau, Costantini, Denissen, Fleeson, Grafton, et al., 2017; Alan, Boneva, and Ertac, 2019; Kosse, Deckers, Pinger, Schildberg-Hörisch, and Falk, 2020; Alan, Baysan, Gumren, and Kubilay, 2021. Fryer and Katz, 2013 elucidates interventions capable of propelling the academic performance and mental health of underprivileged children beyond the confines of poverty's grasp by enhancing school quality. Personality, as evidenced by Robins, Tracy, Trzesniewski, Potter, and Gosling, 2001 and Borghuis, Denissen, Oberski, Sijtsma, Meeus, Branje, Koot, and Bleidorn, 2017, exhibits notable adaptability. Moreover, the role of students' competitiveness in shaping educational and labor market outcomes gains prominence through Buser, Niederle, and Oosterbeek, 2021's research. Meanwhile, individuals actively engage in self-regulated, coordinated efforts to modify their personalities, as explored by Hennecke, Bleidorn, Denissen, and Wood, 2014 and Stieger, Flückiger, Rüegger, Kowatsch, Roberts, and Allemand, 2021. However, not all interventions prove effective. Barrera-Osorio and Linden, 2009 assessed Colombia's Computers for Education program, finding the intervention showed no significant effects on the assessed math and language skills across grades 3 to 9. Similarly, Glewwe, Hanushek, Humpage, and Ravina, 2011 found limited treatment impact on school time and learning from various school and teacher attributes. This paper extends the empirical evidence on student intervention and is of great research significance since it shows that low-cost and accessible interventions could significantly improve students' personality development.

Secondly, this paper contributes to the gap that there is not much evidence on how peer outcomes affect untreated peers who are friends of the treated students and their influencing channels. Most research focuses on how peer characteristics, such as gender, race, or academic achievement, affect outcomes (Hoxby, 2000, Figlio, 2007, Sacerdote, 2001a, Angrist and Lang, 2004, Hoxby and Weingarth, 2005, Lavy and Schlosser, 2011), while studies on peer influence on noncognitive skills remain scarce and debated. Feng, Kim, and Yang (2024) explored the effect of being exposed to left-behind peers on non-cognitive skills, but still they focused on the effect of a peer characteristic, i.e., left-behind status, not peer outcomes, possibly because the endogeneity issue is harder to address in the context of peer outcome effect. Studies on how peer personality affects one's own personality are limited and have lots of disagreement. Shan and Zölitz (2022) found competitive peers increased competitiveness by randomly assigning university students into

different groups and providing each group with a £20 voucher at the university bar to foster peer interactions. In contrast, Bietenbeck (2021) found no significant effect of being exposed to more motivated peers in elementary school, so he suggests that the positive spillover on achievement he found is most likely due to an improved classroom learning environment. Thus, the transferability of personalities is still in the exploratory stage and has lots of disagreement.

The greatest challenge of studying peer effect is that people are self-selected into different groups based on their unobservable characteristics (for example, their hobbies), so it's difficult to distinguish real social effects from correlated effects, i.e., individuals in the same reference group tend to behave similarly because they are alike or face a common environment (Carrell, Sacerdote, and West, 2013, Hsieh and Lee, 2016, Manski, 1993). Evans, Oates, and Schwab, 1992 employs instrumental variables, while Sacerdote, 2001b, Zimmerman, 2003, Whitmore, 2005, Carrell et al., 2013, De Giorgi and Pellizzari, 2014, and Booij, Leuven, and Oosterbeek, 2017 utilize natural experiments to uncover a causal relationship underpinning the influence of peer effects on educational outcomes. Carrell and Hoekstra, 2010 unveils that children exposed to domestic violence negatively impact the test scores of their peers in reading and math. Oosterbeek and Van Ewijk, 2014. List, Momeni, and Zenou, 2020 presents evidence of early childhood interventions affecting the noncognitive skills of both treated and untreated children. The intervention design of this paper effectively addresses the identification challenge by cleanly separating students into treated and untreated groups, creating an exogenous shock on peer outcomes to identify real social effects.

Thirdly, much peer effect research is only able to identify peers in a broad concept like classmates and cohorts due to a lack of network data (Feng et al., 2024, Cattan, Salvanes, and Tominey, 2023). This paper adds new insight to the literature on the identification of peer effect(e.g. Goldsmith-Pinkham and Imbens, 2013, Bramoullé, Djebbari, and Fortin, 2009) and school networks (e.g. Hu, 2023 and Cattan et al., 2023). This study builds on Boucher, Rendall, Ushchev, and Zenou (2024)'s non-linear model of peer effects. However, it differs in three ways. First, this study imposes an exogenous shock through the intervention that better resolves the endogeneity issue compared to Boucher et al. (2024), where they rely on a set of exogenous characteristics that might not fully address the unobservable bias. Additionally, this study uses data from rural Chinese adolescents (2015-2017), contrasting with the predominantly Western focus in the literature. Finally, this paper does not impose a strategic choice in the model like Boucher et al. (2024), Duflo, Dupas, and Kremer (2011), Carrell et al. (2013), and Booij et al. (2017). Theoretical contributions by Calvó-Armengol, Patacchini, and Zenou, 2009 propose a unique Nash equilibrium within the context of peer effects, wherein each participant's strategic choice aligns with their Katz-Bonacich centrality and is elaborated upon through theoretical modeling. However, this is not convincing in a noncognitive skills setting as children are unlikely to choose their resilience or depression status. Shifting towards empirical examination, Bramoullé et al., 2009 discusses the prerequisites and requirements for identifying endogenous and exogenous peer impacts in a network interaction. This study aims to provide a model to identify which friends have the most significant impact on the targeted outcomes of

the intervention when there is an intervention with part of the treatment group being treated and access to their network data.

3 Data and Measurement

3.1 Interventions

3.1.1 The Bedtime Story Project of Growing Home Foundation

The bedtime story project plays bedtime stories to boarding students at night to improve students' sleep conditions and physical and mental health. All interventions were run by the Beijing Growing Home Foundation, a public welfare organization that serves rural boarding schools and rural children's growth and education in China. In the early stages of the project, professional organizations screened out 1,001 high-quality bedtime stories involving "growth healing, character philosophy, character inspiration, knowledge vision, witty adventure, campus students, fairy tales, and myths". Stories were recorded and shared by China Central Radio Station. The Growing Home Foundation purchased high-quality speakers and distributed them to treated schools. Each student dormitory is equipped with a speaker to ensure that students can hear bedtime stories clearly.

A noteworthy aspect of this project lies in its remarkably low marginal cost per student served. This observation is supported by an examination of the financial records of Tanpan Primary School and Malin Primary School in Chongqing province for the year 2017[1]. Specifically, the total expenditure associated with the acquisition of 75 new speakers and their subsequent operational costs amounted to 9,255.76 RMB. This investment facilitated audio bedtime story access for a cohort of 503 boarding students, resulting in an average cost of 18.4 RMB (around two pounds) per student for the inaugural month. Subsequently, with the equipment firmly in place, the ongoing operational expenses were reduced significantly to a mere 10.49 RMB (around 1.15 pounds) per student per month. This efficient cost structure underscores the project's cost-effectiveness in providing valuable services to a wide student population.

From December 2015 to September 2017, we did randomized control trials (RCTs) and collected data to evaluate the interventions. We will keep track of our samples in 2020 and will continue in the future when they enter universities and the labour market. After September 2017, the project expanded from selected schools to all schools and is known as the "New One Thousand and One Nights" public welfare project. By the end of June 2020, the "New One Thousand and One Nights" project will have served 10,039 schools in 948 counties in 29 provinces, bringing companionship to more than 3.99 million children.

[1] See http://www.zhenrogy.org/upload/2018/20181129101056_7665.pdf.

3.2 Experimental Design and Timeline

The project selected 137 rural boarding schools in 5 counties, including Zhuolu County, Guyuan County, and Yu County in Hebei Province, and Cangxi County and Wangcang County in Sichuan Province. The four selection criteria are: first, the counties are national-level poverty-stricken; second, there is a certain scale of boarding primary schools in the county; third, the proportion of ethnic minorities is not high so that the students can understand the stories presented in Mandarin; and finally, relevant departments such as education bureaus and schools in project counties are willing to support project implementation and data collection.

The baseline survey, conducted in October 2015, served as the initial point of reference for this study. In November 2015, we employed random assignment to allocate schools in Cangxi, Yu, Guyuan, and Zhuolu counties into three distinct treatment groups, each accompanied by corresponding control groups. No students received any form of treatment in schools in Wangcang County, constituting a separate control group.

The three distinct treatments employed in this study encompassed the following interventions, as shown in Table 1: 1) the provision of audio bedtime stories exclusively; 2) the provision of storybooks exclusively; 3) the simultaneous provision of both audio bedtime stories and storybooks. It is crucial to clarify that, henceforth, when referring to "treatment" in this paper, we specifically mean the "bedtime-story-only treatment." This distinction is important as the study's primary focus revolves around examining the transfers of personality development associated with this particular intervention.

Table 1. Baseline Distribution of Students by Treatment

Treatment	Schools	Students	%	Boarding Students	%
Audio bedtime story	30	3,911	21.97	2,174	55.59
Books	30	3,673	20.64	2,493	67.89
Audio bedtime story + books	29	3,868	21.73	2,349	60.73
Control (schools)	33	4,325	24.30	2,675	61.85
Control (counties)	15	2,021	11.36	1,281	63.38
Total	137	17,798	100.00	10,972	61.65

Data source: Baseline survey data of the Bedtime Story Intervention Project in 2015.

After the project counties are determined, the sample schools are obtained by simple random sampling, and then two classes of grade 4 and grade 5 are randomly selected in the sample schools, and all students in the sample classes that are drawn automatically become sample students. In this way, sampling is divided into two levels: school sampling and class sampling.

In terms of school sampling, the research team obtained a list of all primary schools in the surveyed counties from the website of the Ministry of Education's student physique reporting system. The list is submitted to the county education bureau to obtain basic information about the school, including whether it is a boarding school, the distance from the

school to the county seat, the number of students in the school, the number of teachers, and the number of boarding students. Non-boarding schools, schools with less than 20 students in grades 4 and 5, and schools with less than 10 boarding students were removed, and the remaining schools formed the school sample frame of the project county.

Then, the research classes were selected in the sample schools according to the principle of random sampling. In principle, there are two survey classes for each grade. When there are only 1 or 2 classes in each grade, all classes in that grade will be included in the survey; if the number of classes in a grade is greater than 2, then it will be used. The method of random number tables randomly selects two classes as the object of investigation. According to this sampling rule, we finally determined that each grade has at least 1 class and a maximum of 2 classes. In principle, all students in the sample class automatically become sample students.

Students who have one of the following situations are not considered as sample students: students are not at school for some reason on the day of the survey; students are unwilling to cooperate with the survey; students with severe intellectual disabilities are unable to participate in the survey; students with severe emotional problems or emotional disorders are unable to complete the questionnaire and take the test.

We track the implementation of the treatment by asking the school prime minister and students about the intervention. The audio bedtime story intervention is defined as "successfully implemented" if there are some students in the school who report "having listened to the stories". Those in the same school who report themselves not having listened may be day students who do not live in a school dormitory or students who forgot for personal reasons. The storybooks are provided in the corner of each classroom, so students have full access to them at school. They can also borrow and bring the books to the dormitory or home by applying.

Figure 1. Survey and Treatment Timeline

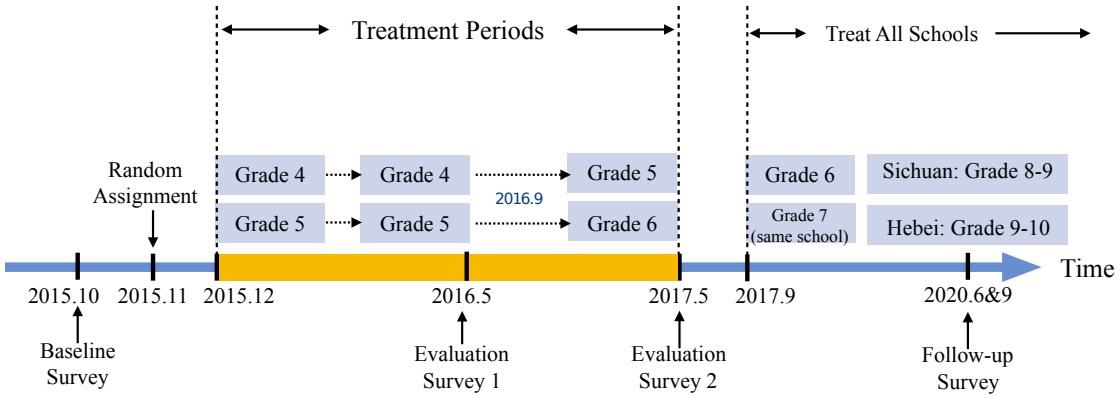


Figure 1 shows the timeline for the whole treatment and data collection process. The intervention commenced in

December 2015 when the students were in grades 4 and 5, with the initial follow-up surveys occurring in May 2016, corresponding to the same academic year as 2015. In essence, the interventions spanned a six-month period. However, accounting for the Chinese winter vacation, which typically extends from mid-January to late February, and the intervention's late December commencement, the effective intervention duration is four months. The second follow-up survey took place in May 2017, coinciding with the students' transition to grades 5 and 6. Similarly, factoring in summer (June to August) and winter breaks, the intervention period was reduced from 18 months to 14 months.

In 2020, the surveys in Sichuan and Hebei provinces took place in different months because of COVID-19 travel restrictions. Surveys in Sichuan province were in June, whereas surveys in Hebei province were in September and November. Since the new academic year starts in September, the students were in the 2019 academic year and mainly in grades 8 and 9, which means they were all in junior high schools. In Hebei province, surveys were conducted in September, November, and January, so the students were in the 2020 academic year and mainly in grades 9 and 10, so some of them are already in senior high schools.

4 Measurement

4.1 Friendship Network

We adopted the data collection approach used in the National Longitudinal Study of Adolescent to Adult Health (Add Health) to gather friendship information for each student in the study at both baseline and follow-up. Specifically, we asked students to list up to five closest friends within their classroom and rank them according to their level of closeness.

4.2 Non-cognitive Skills and Mental Health

There are two methods for measuring non-cognitive skills, self-reported psychological scales completed by students and behavioural observations during questionnaire completion. In the initial three phases of the research, the research team employed resilience, internalizing and externalizing behaviours, and the Children's Depression Inventory as measures. In the fourth phase of the study, the research team introduced the Big Five Personality Inventory, the Short Grit Scale, and the Short Self-Control Scale in consideration of the potential measurement errors associated with self-reported psychological scales used for assessing non-cognitive skills. Consequently, it is intended to utilize students' observed behaviours while completing questionnaires as an alternative means to more accurately gauge their non-cognitive skill levels.

We explain the measure of resilience for example. It is defined as the capacity of a child to withstand and overcome adversity. The resilience scale was adapted from the "Gansu Basic Education Tracking Survey Project", a long-term study involving prominent international researchers such as Albert Park, Emily Hanuman, and Wang Jiayi. This scale is based on Song (2003) and consists of 39 questions posed to the responding student. It evaluates various dimensions of resilience, including optimism, self-efficacy, relationships with adults and peers, interpersonal sensitivity, and emotional control. These dimensions collectively assess a child's ability to remain positive, believe in their capabilities, navigate adult and peer interactions, exhibit emotional control, and effectively manage interpersonal dynamics. The explanations for other non-cognitive outcomes are shown in subsection 11.1.

It is imperative to note that all the outcomes under consideration were predetermined unless explicitly specified otherwise. Measurements are based on survey questions in past literature measuring similar outcomes and are in subsection 11.1.

4.2.1 Exploratory Factor Analysis

Non-cognitive skills in childhood are inherently challenging to measure. As discussed earlier, various measures assess multiple dimensions of a child's non-cognitive abilities. Following Attanasio, de Paula, and Toppeta, 2024, we utilize exploratory factor analysis to determine the relevance of each measure to specific factors.

The EFA is based on decomposing the polychoric correlation matrix of the items and using maximum likelihood estimation (Olsson, 1979)[2]. Table 3, Table 4, and Table 5 present the estimates of EFA. The factor loadings in this Table show a clear separation between items. In Table 3 and Table 4, factor 1 explains interpersonal resilience, factor 2 explains personal resilience, factor 5 explains self-esteem questions, and factor 4 explains depression questions. factor 7 in Table 5 explains internalization questions and factors 1-3 explain externalization questions. We highlight that the factor loadings have a similar magnitude across groups, pointing out that there is a similar association between the item and the factor across groups,Äîthe suggested number of factors to retain is shown in Table 2.

Table 2. Suggested number of factors to retain

Approach	All
Parallel analysis	4
Kaiser	4
Velicer MAP	4

Notes: The table compares the optimal number of factors for all questions related to resilience, self-esteem, internalization, and externalization suggested by different approaches for the novel scale: parallel analysis (Horn (1965)), Kaiser's eigenvalue rule (Kaiser (1960)), Velicer Minimum Average Partial test (MAP, Velicer and Jackson (1990)).

[2]The polychoric correlation is an estimate for the correlation between two normally distributed continuous random variables observed as ordinal variables. The solution is rescaled using oblique factor rotation obtained via the PROMAX protocol outlined in Hendrickson and White (1964) (with k = 3). Since we use a dedicated factor structure based on the oblique factor rotation matrix suggested above, our factor scores (i.e., skills) are not orthogonal. We use the Python function FactorAnalyzer with rotation='oblimin', method='minres'.

4.2.2 Factor model

We specify a factor model to estimate the relationship between internalizing and externalizing skills and the available measures, the Rutter and SDQ items, based on the results of the exploratory analysis described in section subsubsection 4.2.1. Each individual is denoted by $i = 1, ..., N$. For each individual, we observe categorical item responses Z_{ij}, corresponding to the survey questions (subsection 11.1) where j indexes the 86 available items. We follow the literature where our survey questions for resilience come from and assume a latent bi-dimensional vector of resilience $Yi = (Y_i^{Personal-Resilience}, Y_i^{Social-Resilience})$(Song, 2003), Other socio-emotional skills $Yi = (Y_i^{Self-esteem}, Y_i^{Externalization}, Y_i^{Internalization})$ and mental health $Y_i^{Depression}$.

The relationship between the latent factors Yi and the available measures Z_{ij} is characterized by item- and group-specific intercepts v_j and loadings ł$_j$ and is affected by an independent measurement error term u_{ij}. The measures are defined in terms of the following variable:

$$Z_{ij}^* = v_j + \lambda_j^T Y_i + u_{ij}.$$

We consider a dedicated factor structure, where each item loads only on one latent dimension, and follow the structure found in the exploratory factor analysis (Heckman, Pinto, and Savelyev, 2013). We estimate a baseline model which is characterized by the bare minimum number of assumptions with the parameterisation defined below (Wu and Estabrook, 2016).

Given the specification of the behavioural scale, the measures Z_{ij} have a discrete nature and take one of an ordered number of values. To allow for such measures, we introduce item-specific threshold parameters τ_j as follows:

$$Zij = s \;\; if \;\; \tau_{s,j} \geq Z_{ij}^* \geq \tau_{s+1,j} \;\; for \;\; s = 1,2,3,4$$

with $\tau_{0,j} = -\infty$ and $\tau_{3,j} = \infty$. The values taken by Z_{ij}, $s = 1,2,3,4$, correspond to those in our data ("Strongly disagree" = 1, "Disagree" = 2, "Agree" = 3, "Strongly Agree" = 4)[3].

Following the literature, we assume that the latent factors and the measurement error terms are normally distributed:

$$Y_i \sim N(\kappa, \sigma_Y^2) \;\; and \;\; u_{ij} \sim N(0, \sigma^2).$$

Finally, we make the normalization assumption needed to deal with factor indeterminacy by setting the mean κ and the variance σ_Y^2 of the factor equal to 0 and 1 respectively. In addition, the intercepts v_j are equal to zero and the error variance σ^2 to 1, while the loadings λ_j and threshold τ_j are free to vary.

[3] We employed an ordered logit model to estimate the latent continuous variable underlying the ordinal observed variable using the OrderedModel function from the statsmodels library.

Table 3. Loadings from exploratory factor analysis for resilience, self-esteem, and depression

	factor 1	factor 2	factor 3	factor 4	factor 5	factor 6
Resilience						
Peer Relationship	**0.355**	0.052	0.199	0.039	-0.041	-0.396
Adult Relationship	0.345	**0.443**	0.267	-0.009	0.032	-0.083
Self-Efficacy	0.224	**0.467**	-0.069	-0.086	0.070	0.379
Interpersonal Sensitivity	0.309	-0.024	**0.391**	-0.002	0.065	-0.456
Optimism	**0.493**	0.000	0.086	-0.009	0.027	-0.575
Interpersonal Sensitivity	-0.367	**0.385**	0.133	-0.250	0.166	0.253
Optimism	**0.850**	0.222	-0.107	-0.043	-0.031	0.022
Self-Efficacy	**0.449**	0.081	0.397	-0.060	0.077	0.002
Peer Relationship	**0.864**	0.076	0.011	-0.046	0.029	0.085
Adult Relationship	0.108	**0.818**	-0.068	-0.054	0.022	0.049
Self-Efficacy	**0.781**	0.078	-0.014	0.053	-0.008	-0.128
Interpersonal Sensitivity	**0.883**	-0.043	-0.004	-0.045	0.064	-0.125
Emotional Control	0.268	**0.665**	-0.164	-0.048	0.041	0.184
Peer Relationship	-0.098	**0.369**	0.082	-0.205	0.173	0.419
Adult Relationship	**0.632**	0.384	0.052	0.032	-0.023	-0.167
Self-Efficacy	0.065	**0.556**	0.024	-0.103	0.088	0.344
Self-Efficacy	-0.082	**0.510**	0.119	-0.201	0.140	0.291
Peer Relationship	**0.896**	0.022	0.009	-0.032	0.009	-0.021
Optimism	**0.501**	0.174	0.335	0.013	0.066	-0.112
Optimism	0.049	**0.809**	0.004	-0.058	0.064	0.053
Emotional Control	**0.679**	0.043	0.212	-0.025	0.095	-0.078
Adult Relationship	0.045	**0.558**	0.035	-0.137	0.127	0.298
Self-Efficacy	0.070	**0.414**	-0.003	-0.150	0.110	0.410
Optimism	0.047	**0.600**	-0.170	-0.097	0.031	0.258
Adult Relationship	**0.496**	0.037	0.409	-0.016	0.058	0.020
Emotional Control	0.071	**0.565**	0.180	-0.100	0.136	0.199
Self-Efficacy	0.323	0.099	**0.562**	-0.046	-0.004	-0.055
Optimism	0.157	0.237	-0.017	-0.173	0.154	**0.537**
Interpersonal Sensitivity	0.179	0.342	-0.074	-0.175	0.131	**0.431**
Optimism	**0.655**	-0.162	0.303	-0.080	0.066	0.144
Peer Relationship	0.157	**0.508**	-0.237	-0.134	0.043	0.348
Self-Efficacy	0.154	-0.134	**0.294**	0.037	0.013	-0.587
Adult Relationship	0.208	0.150	**0.641**	0.018	0.082	-0.070
Emotional Control	**0.475**	-0.005	0.450	-0.008	0.004	-0.060
Peer Relationship	-0.296	0.294	-0.092	-0.194	0.148	**0.406**
Optimism	-0.067	**0.800**	0.110	-0.049	0.105	0.041
Adult Relationship	0.459	-0.037	**0.491**	-0.043	0.048	0.199
Self-Efficacy	**0.666**	0.000	0.223	-0.037	0.041	0.228
Peer Relationship	0.023	**0.738**	0.113	-0.071	0.106	0.065
Optimism	**0.719**	0.080	0.130	-0.027	-0.020	0.236
Emotional Control	**0.697**	-0.038	0.150	-0.018	0.062	-0.037
Adult Relationship	**0.449**	0.404	0.255	-0.006	-0.012	-0.033
Peer Relationship	**0.449**	0.404	0.255	-0.006	-0.012	-0.033
Interpersonal Sensitivity	**0.646**	-0.049	0.196	-0.008	0.018	-0.101

Note: The table displays the factors loadings obtained from exploratory factor analysis (EFA) by sample. The EFA is based on the decomposition of the polychoric correlation matrix, and uses PROMAX rotation. Since they are all behaviours indicating lower skills, we recode all of them in reverse, i.e. "Strongly disagree" = 1, "Disagree" = 2, "Agree" = 3, "Agree wholeheartedly" = 4.

5 Treatment Effect Evaluation

5.1 Estimation

The empirical strategy is based on the following estimating equation:

Table 4. Loadings from exploratory factor analysis for resilience, self-esteem, and depression (Cont.)

	factor 1	factor 2	factor 3	factor 4	factor 5	factor 6
Self-esteem 1	0.217	0.166	**0.670**	-0.062	0.015	0.005
Self-esteem 2	-0.039	-0.124	-0.088	-0.010	**0.968**	0.033
Self-esteem 3	-0.098	-0.055	**0.949**	-0.037	0.037	-0.103
Self-esteem 4	0.198	-0.013	**0.762**	-0.034	-0.013	0.036
Self-esteem 5	-0.212	-0.030	0.026	-0.064	**0.574**	0.050
Self-esteem 6	-0.017	-0.089	-0.107	0.033	**1.062**	-0.030
Self-esteem 7	0.161	0.037	**0.739**	-0.008	0.035	-0.062
Self-esteem 8	**0.477**	0.215	0.355	0.000	-0.081	-0.098
Self-esteem 9	-0.024	0.099	0.019	0.037	**0.881**	-0.168
Self-esteem 10	**0.369**	0.046	0.532	-0.007	0.011	-0.062
Depression 1	-0.137	-0.139	-0.121	**0.500**	-0.190	-0.267
Depression 2	-0.135	-0.124	-0.150	**0.547**	-0.178	-0.184
Depression 3	0.051	-0.054	-0.086	**0.699**	-0.136	-0.027
Depression 4	-0.205	-0.319	-0.473	-0.471	-0.221	**-0.035**
Depression 5	-0.243	-0.159	-0.011	**0.518**	-0.156	-0.190
Depression 6	-0.066	-0.161	-0.239	**0.493**	-0.223	-0.223
Depression 7	-0.158	-0.128	-0.143	**0.531**	-0.209	-0.242
Depression 8	**-0.020**	-0.069	-0.450	-0.731	-0.137	-0.084
Depression 9	-0.158	-0.254	-0.114	**0.500**	-0.185	-0.132
Depression 10	-0.110	-0.286	-0.172	**0.499**	-0.201	-0.014
Depression 11	-0.158	-0.253	-0.152	**0.436**	-0.181	-0.141
Depression 12	-0.345	-0.396	-0.282	-0.324	**-0.192**	-0.181
Depression 13	**0.232**	-0.191	0.138	0.542	0.001	0.156
Depression 14	-0.056	-0.610	-0.172	**0.286**	-0.149	0.091
Depression 15	-0.154	-0.656	-0.091	**0.275**	-0.106	0.184
Depression 16	-0.296	-0.457	-0.296	-0.191	**-0.172**	-0.130
Depression 17	-0.057	-0.329	-0.226	**0.432**	-0.198	-0.069
Depression 18	-0.075	-0.380	-0.216	**0.401**	-0.194	-0.056
Depression 19	-0.081	-0.524	-0.213	**0.310**	-0.169	0.049
Depression 20	-0.086	-0.394	-0.159	**0.412**	-0.209	0.060

Note: The table displays the factors loadings obtained from exploratory factor analysis (EFA) by sample. The EFA is based on the decomposition of the polychoric correlation matrix, and uses PROMAX rotation. Since they are all behaviours indicating lower skills, we recode all of them in reverse, i.e. "Strongly disagree" = 1, "Disagree" = 2, "Agree" = 3, and "Agree wholeheartedly" = 4.

$$Y_{is,t} = \beta_0^k + \beta_1^k T_{s,0} + \sigma^k Y_{is,0} + \gamma^k X_{is,0} + \alpha_c^k + \varepsilon_{is,t}^k, \tag{1}$$

where k $\in$ {day students, board students}

where $Y_{is,t}$ represents the outcome for individual i in school s at time t, and $T_{s,0}$ is a binary treatment indicator that equals 1 if school s is assigned to the treatment group. The baseline outcome, $Y_{is,0}$, is included to control for pre-treatment differences. Control variables $X_{is,0}$ include baseline characteristics such as parental education level and other imbalanced covariates between the treatment and control groups. These baseline variables encompass essential attributes, including age, age squared, gender, grade, baseline weight, baseline height, father's years of education,

Table 5. Loadings from exploratory factor analysis for internalization and externalization

	factor 1	factor 2	factor 3	factor 4	factor 5	factor 6	factor 7	factor 8
INT1	-0.214	0.006	-0.392	0.339	0.143	-0.032	-0.128	**0.176**
INT2	0.074	-0.208	-0.259	0.076	0.121	0.186	**0.240**	0.180
INT3	-0.139	-0.001	-0.733	-0.097	0.084	-0.084	**0.064**	-0.108
INT4	**0.020**	-0.076	-0.052	-0.053	-0.842	-0.038	-0.011	-0.103
INT5	0.028	-0.292	0.257	-0.035	-0.067	0.068	**0.186**	0.215
INT6	-0.009	-0.220	0.234	0.031	0.047	-0.045	0.084	**0.539**
INT7	-0.107	**0.002**	0.191	-0.068	-0.712	-0.039	0.020	0.119
INT8	-0.124	0.031	**0.393**	0.089	0.205	-0.030	0.012	0.091
INT9	0.055	-0.298	**0.294**	-0.197	0.323	-0.214	-0.194	-0.283
EXT1	-0.206	**0.217**	-0.163	0.036	0.015	0.393	-0.174	-0.086
EXT2	-0.209	0.116	0.120	-0.041	0.032	**0.688**	-0.138	0.035
EXT3	0.361	0.084	0.010	-0.122	0.094	0.036	-0.212	**0.474**
EXT4	0.066	**0.754**	0.005	-0.036	0.043	0.028	0.061	-0.015
EXT5	0.055	**0.620**	0.005	-0.044	-0.010	0.049	-0.037	0.077
EXT6	0.009	**0.330**	0.034	-0.175	0.036	-0.046	-0.356	0.336
EXT7	-0.046	**0.635**	-0.022	-0.057	0.052	0.042	-0.067	-0.098
EXT8	-0.010	-0.036	0.019	**0.880**	0.049	-0.011	-0.040	-0.017
EXT9	0.047	-0.012	0.045	**0.938**	0.010	-0.023	0.024	-0.012
EXT10	0.174	-0.227	0.017	0.174	0.077	-0.026	**0.372**	0.163
EXT11	0.062	-0.072	0.023	-0.071	0.010	-0.130	**0.518**	0.020
EXT12	0.132	-0.237	**0.447**	-0.007	0.100	-0.105	0.236	0.121
EXT13	-0.069	0.182	0.079	-0.067	-0.007	-0.135	**0.532**	-0.091
EXT14	0.141	-0.110	**0.466**	0.043	0.159	-0.083	0.292	0.059
EXT15	0.170	-0.204	**0.426**	0.010	0.186	-0.089	0.207	0.042
EXT16	0.033	-0.008	**0.385**	-0.019	0.246	-0.029	0.224	-0.047
EXT17	0.150	-0.172	**0.290**	-0.080	0.276	-0.014	0.103	-0.074
EXT18	0.227	-0.094	0.006	-0.020	0.041	**0.736**	0.052	-0.042
EXT19	**0.387**	0.108	0.009	-0.001	0.096	0.319	0.124	0.133
EXT20	**0.733**	-0.066	-0.046	0.130	0.039	0.102	0.046	0.023
EXT21	**0.769**	-0.059	0.078	-0.016	0.051	-0.079	0.030	0.053
EXT22	**0.637**	0.185	-0.059	-0.021	0.085	0.041	0.151	0.057
EXT23	0.572	0.194	0.027	**0.328**	-0.035	0.098	-0.087	-0.029
EXT24	**0.659**	0.013	0.107	0.026	0.064	0.059	-0.024	0.042
EXT25	**0.688**	-0.069	0.203	-0.013	0.069	-0.082	-0.025	-0.044

Note: The table displays the factors loadings obtained from exploratory factor analysis (EFA) by sample. The EFA is based on the decomposition of the polychoric correlation matrix and uses PROMAX rotation. Since they are all behaviors indicating lower skills, we recode all of them in reverse, i.e. "Strongly disagree" = 1, "Disagree" = 2, "Agree" = 3, and "Agree wholeheartedly" = 4.

mother's years of education. County-level fixed effects are captured by α_c, and standard errors are clustered at the school level. The regression is run separately for boarding and day students to capture the treatment effect and the spillover effect.

The fundamental assumption that is crucial to our methodology is the effectiveness of our randomization process. To validate this assumption, we examine whether a student's baseline characteristics are indicative of their assignment to either the treatment or control group. We assess this with a balance test, with the dependent variable being a binary indicator that represents whether the student belongs to the treatment group or the control group.

As detailed in the referenced work, there is no discernible indication of differential attrition associated with treatment

status. Those who do not experience attrition exhibit a harmonious distribution of baseline characteristics. Attrition is also not related to baseline outcome variables like resilience, self-esteem, etc.

5.2 Validation for the RCT

5.2.1 Balance Test

The samples this paper focuses on are the "audio bedtime story only" treatment group and the control group. Since the speakers worked in the school dormitories, only board students in the treatment group were directly treated. The day students, in contrast, were only affected through spillover and peer effects, if any. Table 6 presents the balance test for key characteristics for the board students and day students sample in the "audio bedtime story only" treatment group and the control group, and the balance test for the full sample is in section 11. All results are well-balanced.

Table 6. Balance Test for Board Students and Day Students

	Board Student			Day Student		
	Control	Treat	P-value/ RW p-value	Control	Treat	P-value/ RW p-value
Personal Characteristics:						
Age	10.316	10.266	0.610	10.146	10.108	0.757
	(0.828)	(0.866)	[0.979]	(0.798)	(0.792)	[0.985]
Gender	0.529	0.517	0.550	0.443	0.499	0.004
	(0.499)	(0.500)	[0.997]	(0.497)	(0.500)	[0.719]
Grade	4.527	4.512	0.650	4.461	4.456	0.844
	(0.499)	(0.500)	[0.996]	(0.499)	(0.498)	[1.000]
Height	138.912	138.810	0.912	139.078	138.316	0.387
	(7.701)	(7.885)	[1.000]	(7.740)	(7.606)	[0.823]
Weight	33.367	32.844	0.528	33.555	32.911	0.521
	(8.293)	(7.344)	[0.960]	(8.635)	(7.489)	[0.873]
Left-behind	0.430	0.456	0.627	0.408	0.423	0.788
	(0.495)	(0.498)	[0.981]	(0.492)	(0.494)	[0.989]
Family Characteristics:						
Mother's Education	8.137	8.111	0.837	8.587	8.794	0.213
	(2.004)	(2.067)	[1.000]	(2.118)	(2.201)	[0.823]
Father's Education	8.605	8.618	0.906	8.998	9.370	0.026
	(2.013)	(2.027)	[1.000]	(2.192)	(2.187)	[0.565]
Parents' Marriage	0.134	0.130	0.087	0.098	0.440	0.465
	(0.341)	(0.337)	[1.000]	(0.281)	(0.297)	[0.989]
Outcome Variables:						
Resilience	-0.015	0.002	0.780	0.046	0.046	0.999
	(0.996)	(0.956)	[1.000]	(1.003)	(0.971)	[1.000]
Self-esteem	-0.374	-0.339	0.427	-0.382	-0.376	0.883
	(0.909)	(0.865)	[0.996]	(0.866)	(0.897)	[1.000]
Sleeping quality	35.859	35.785	0.746	34.871	34.782	0.792
	(5.120)	(4.929)	[1.000]	(5.302)	(5.137)	[1.000]
Depression	0.024	0.074	0.417	-0.114	-0.127	0.858
	(1.000)	(0.983)	[0.981]	(0.978)	(0.958)	[1.000]
Math	0.185	0.125	0.100	0.154	0.376	0.025
	(0.827)	(0.804)	[0.960]	(0.815)	(0.709)	[0.193]
Reading	-0.402	-0.428	0.750	-0.435	0.022	0.383
	(0.816)	(0.779)	[0.996]	(0.807)	(0.757)	[0.849]
# friends	4.030	3.795	0.482	4.115	3.914	0.398
	(1.362)	(1.695)	[0.689]	(1.256)	(1.573)	[0.713]

Note: This table presents the means and standard deviations (in parentheses) by treatment group. P-values test the difference between each treatment group's mean and the control. P-values are two-sided. RW P-values use the Romano-Wolf method to account for multiple hypothesis testing across all outcome measures in the table. Outcome variables are standardized factor scores for noncognitive skills and depression.

5.2.2 Attrition and Estimating Power

According to the survey data, in general, 560 students from the baseline survey in October 2015 did not appear in the May 2016 survey, an attrition rate of 3.3%. The 1,203 students surveyed in the October 2015 survey did not appear in the May 2017 survey, an attrition rate of 7.1%. In 2020, 4,927 students from the baseline survey did not appear, an attrition rate 27.7% of the in-person survey. Aside from surveying the students in person, we also track them by sending them online surveys with a shorter list of questions than the in-person survey, getting information by telephone, and checking their records at the local education bureau. If we take other surveying sources into consideration, the attrition rate in 2020 would be 15.3%. The attrition rates in 2016 and 2017 are comparably small to similar studies with primary school students, which report attrition rates of 20.6 percent (Muralidharan and Sundararaman, 2011), 16.1 percent (Kosse et al., 2020) and 20 percent(Leaver, Ozier, Serneels, and Zeitlin, 2021).

In addition, there were new participants in the project during its implementation. In the 2016 follow-up survey, 640 students were newly enrolled, with a new participation rate of 3.8%, and in 2017, there were 390 new entrants, with a participation rate of 2.3%. In 2020, there were 4,884 new entrants, with a participation rate of 27.3%. The new participants are designed for the replacement of missing samples to maintain the scale of the total sample.

Nevertheless, it is imperative to acknowledge that the attrition rate recorded in the year 2020 remains non-negligible. To address this attrition and ensure statistically robust analyses, a recommended sample size was calculated employing the power analysis software Optimal Design(Spybrook, Bloom, Congdon, Hill, Martinez, Raudenbush, and TO, 2011). Under the assumption of a power level of 0.80 and an effect size of 0.1, a total sample size of 3,170 individuals across the two groups would be requisite. For a power level of 0.80 and an effect size of 0.1 and an effect size of 0.2, 800 would be requisite. Therefore, our working sample of 3,919 students per wave who showed up in all four surveys is enough for analysis.

Table 7. Attrition

	Board Student	Day Student	Board Student	Day Student
2016	0.244	0.758	0.154	0.569
	[0.284]	[0.787]	[0.295]	[0.823]
2017	0.000	0.138	0.000	0.054
	[0.001]	[0.333]	[0.001]	[0.184]
2020	0.151	0.368	0.531	0.713
	[0.284]	[0.625]	[0.504]	[0.823]
Control	No	No	Yes	Yes

Note: Table shows the p-value and Romano-wolf p-value of whether attrition is significantly different across the treatment group and control group.

Table 7 shows that attrition is not significantly different across the treatment group and the control group for board students in 2016 and 2020 and for day students in all rounds. Therefore, the spillover effects in 2016 and 2017, treatment effects in 2016, and all results in 2020 do not pose concerns related to attrition. Figure 2 presents that

attrition is also not related to baseline outcomes.

	Attrition (Day students)	Attrition (Board students)
	(1)	(2)
Cohort's Personal Resilience	-0.017	0.005
	(0.031)	(0.024)
Cohort's Social Resilience	0.016	0.006
	(0.022)	(0.018)
Cohort's Internalization	-0.010	0.003
	(0.007)	(0.008)
Cohort's Externalization	0.007	0.002
	(0.008)	(0.008)
Cohort's Self-esteem	0.011	-0.001
	(0.013)	(0.011)
Cohort's Depression	0.007	0.008
	(0.005)	(0.006)
Observations	2,374	3,342
County FE	Yes	Yes
Other controls	Yes	Yes

Note: This table examines the relationship between attrition and baseline outcomes, including county fixed effects and control variables. Attrition is a dummy variable that takes value 1 if the student was surveyed at baseline but not in 2017 follow-up survey. Standard errors are clustered at the school level. The results indicate that attrition is not correlated with baseline outcomes.

Figure 2. Attrition and Baseline Outcomes

We also observe that there is no statistically significant distinction between the treatment and control groups in terms of the mean student attributes (including age, gender, and family characteristics) among the boarding students who participated in both the 2015 baseline survey and the 2017 follow-up survey (see Table 11). So even though attrition is not balanced between the treatment group and the control group, the non-attritors are balanced in terms of observed characteristics.

5.2.3 Double-blind

A distinct separation of responsibilities was adopted within the research protocol in order to prevent potential bias problems and guarantee the integrity of the investigation. First off, the survey team, entrusted with carrying out the surveys, operated independently from the research team that was in charge of questionnaire design and data processing. Second, the implementation of the intervention was run by the Beijing Growing Home Foundation and the schools, which were kept apart by the research team. The treatment team was entirely responsible for the consistent assignment of the intervention, whereas the survey team handled all interviews for the baseline and follow-up surveys. Thirdly, the survey teams for each round were independent of one another to avoid any prejudice resulting from their prior interactions with the students. Notably, all teams received intensive training under the supervision of the research

team. With the double-blind method used in this design, experiment participants' names were kept secret from those conducting data collection and analysis. Those who were compiling the data simultaneously were unaware of whether each respondent had received the treatment. This technique efficiently minimized potential sources of bias, protecting the experiment's overall integrity (Banerjee and Duflo, 2007).

5.2.4 Hawthorne Effect and Spill-over

People may alter their actions and reactions when they are aware that they are taking part in an experiment (Franke and Kaul, 1978). To address this concern, we purposefully kept the treatment group and the control group unaware of the existence of this comparison or of one another. This precaution serves the twin purposes of preventing contamination and spillover effects as well as making it easier to determine the results of the intervention. However, we are aware of how challenging it is to completely allay the worry of potential spill-over effects if the children have personal contacts from other institutions. So we also selected Wangcang County's schools as a unique control group we refer to as the "Control County." No pupils got any kind of treatment in the "Control County," minimizing the chance of any spillover effects, should they arise.

5.2.5 Selective Migration

Certain experimental investigations, such as those cited in the studies Banerjee, Duflo, Glennerster, and Kinnan (2015) and Nakano and Magezi (2020), may raise questions about selective migration, especially when the intervention incorporates bonuses or credits. Selective migration alludes to the possibility that respondents might decide to move to regions where the intervention is implemented in order to take advantage of the benefits connected with it, thereby adding bias into the statistical results. Nevertheless, in our particular situation, the risk of selective migration was judged to be low, mostly because our intervention was non-monetary and a double-blind approach was strictly followed.

5.2.6 Naturalness

Our study was conducted in a cluster of rural primary boarding schools where peer interaction was a common occurrence. Both the initial and final surveys were presented as standard student surveys and distributed to all students in the classroom, irrespective of whether they were part of the treatment group or not. Consequently, the participants had no knowledge that their self-reported personality scores would be employed in the analysis of personality development and the influence of peers. The variations we investigate in this paper stem from natural fluctuations in group composition, similar to what would naturally happen in various other settings. In essence, we conducted a field experiment that was naturally designed, free from artificial interventions.

5.3 Multiple Hypothesis Testing

We use the step-down method introduced by Romano and Wolf (2005a) and Romano and Wolf (2005b) to get around the problem associated with multiple hypothesis testing, where the risk of incorrectly rejecting at least one null hypothesis across all measurements of a given outcome can significantly surpass the specified significance level. By taking into account the dependency within the joint distribution of test statistics across numerous hypothesis tests, this approach is intended to asymptotically control the familywise error (FWE) rate, which represents the chance of mistakenly rejecting at least one null hypothesis. We compute corrected p-values for step-down multiple testing utilizing the procedure outlined in Romano and Wolf (2016).

6 Motivating Evidence

6.1 Average Treatment and Spillover Effect

The intent-to-treat (ITT) analysis assesses the direct effects of the intervention on board student outcomes. Figure 3 presents the estimated treatment effects across several outcome measures. In 2016, after four months of intervention, there was a 0.085 standard deviation increase in depression levels among students. This short-term increase may be attributed to the content of the stories, which likely triggered feelings of homesickness. However, this effect diminished by 2017, disappearing after 14 months of treatment, suggesting that the initial emotional response was temporary. There is a 0.086 SD increase in personal resilience after 14 months of treatment.

Note that doing the regression and factor analysis separately induces measurement error. In future research, we will follow Attanasio, De Paula, and Toppeta (2024) to do a joint estimation of regressions and the factor model. At the current stage, we provide overall score results in the appendix as a robust checkThe results using the overall score are shown in subsection 11.2.

	Personal Resilience	Social Resilience	Internalization	Externalization	Self-esteem	Depression
			2016			
Treatment	0.038	0.034	-0.032	-0.015	0.044	0.085**
	(0.053)	(0.053)	(0.045)	(0.037)	(0.050)	(0.037)
P-value	0.479	0.519	0.474	0.695	0.378	0.024
RW p-value	0.847	0.847	0.847	0.847	0.847	0.461
Control Mean	-0.002	-0.003	-0.158	-0.158	-0.014	0.040
Observations	5,055	5,055	4,420	4,488	5,055	4,997
			2017			
Treatment	0.086*	0.074	-0.016	-0.040	0.068	0.004
	(0.049)	(0.048)	(0.054)	(0.052)	(0.047)	(0.040)
P-value	0.084	0.124	0.765	0.444	0.153	0.929
RW p-value	0.373	0.489	0.914	0.847	0.597	0.981
Control Mean	0.043	0.054	-0.112	-0.127	0.041	0.055
Observations	5,084	5,084	4,812	4,836	5,084	5,074

Note: Table presents OLS coefficients for the bedtime story treatment group compared to the control group, with standard errors in parentheses. Dependent variables are factor scores. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. RW P values use the Romano-Wolf method to account for multiple hypothesis testing for all outcome measures in this chapter. Estimates control for all control variables in the regression section of the empirical strategy. The regressions are clustered at the school level.

Figure 3. Intent-to-Treat (ITT) Effect of Treatment on Student Outcomes

In addition to the direct effects on treated students, we examine potential spillover effects on non-treated day students. Figure 4 displays the spillover effects observed among day students who were not directly treated. In 2016, the increase in depression did not have a spillover effect on the untreated day students. In 2017, there was a 0.117 SD increase in day students' personal resilience and a 0.159 SD increase in their self-esteem.

	Personal Resilience	Social Resilience	Internalization	Externalization	Self-esteem	Depression
			2016			
Treatment	0.053	0.044	-0.079	-0.068	0.059	-0.015
	(0.056)	(0.054)	(0.063)	(0.050)	(0.055)	(0.051)
P-value	0.340	0.424	0.216	0.177	0.289	0.775
RW p-value	0.712	0.773	0.544	0.557	0.682	0.912
Control Mean	0.092	0.094	-0.259	-0.248	0.077	-0.123
Observations	3,120	3,120	2,729	2,795	3,120	3,093
			2017			
Treatment	0.117*	0.097	0.019	0.046	0.159***	-0.089
	(0.063)	(0.065)	(0.053)	(0.040)	(0.058)	(0.066)
P-value	0.068	0.140	0.727	0.260	0.008	0.187
RW p-value	0.290	0.476	0.912	0.728	0.076	0.515
Control Mean	0.092	0.104	-0.189	-0.243	0.077	-0.072
Observations	2,894	2,894	2,710	2,763	2,894	2,883

Note: Table presents OLS coefficients for the bedtime story treatment group compared to the control group, with standard errors in parentheses. Dependent variables are factor scores. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. RW P values use the Romano-Wolf method to account for multiple hypothesis testing for all outcome measures in this chapter. Estimates control for all control variables in the regression section of the empirical strategy. Standard errors are clustered at the school level.

Figure 4. Spillover Effects on Non-Treated Students

7 Mechanism

It is noteworthy that the average spillover effect on personal resilience and self-esteem exceeds the average direct treatment effect. Two potential explanations for this are worth considering. First, the treatment effects on board students, who interact with and influence day students, may be stronger than those experienced by other groups.

7.1 Social Networks

The availability of network data allows us to examine heterogeneous treatment effects across different social networks. Figure 5 shows a 0.115 standard deviation increase in personal resilience and a 0.09 standard deviation increase in self-esteem among board students with day-student friends. These effects are significantly larger compared to the negligible increases of 0.048 and 0.034 standard deviations in personal resilience and self-esteem, respectively, among board students without day-student friends. Similarly, board students with social connections exhibit a larger treatment effect compared to those without such connections.

A key mechanism behind the observed spillover effects is the influence of peer networks. The analysis highlights significant spillover effects, particularly among students who share social networks with treated peers. Figure 6 illustrates these findings by comparing the outcomes for day students with and without connections to boarding students. The results indicate that day students with social ties to treated boarders experience stronger treatment effects, with a 0.109 standard deviation increase in personal resilience and a 0.167 standard deviation increase in self-esteem. In contrast, day students without such connections show smaller, statistically insignificant gains of 0.085 and 0.078 standard deviations in personal resilience and self-esteem, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Personal Resilience	Self-esteem	Personal Resilience	Self-esteem	Personal Resilience	Self-esteem	Personal Resilience	Self-esteem
Treatment	0.115*	0.092	0.048	0.034	0.085*	0.064	-0.145	0.014
	(0.060)	(0.056)	(0.059)	(0.065)	(0.049)	(0.048)	(0.212)	(0.191)
P value	0.060	0.104	0.417	0.599	0.087	0.189	0.498	0.944
RW p-value	0.165	0.312	0.618	0.751	0.120	0.312	0.751	0.942
Observations	2,389	2,389	2,695	2,695	4,947	4,947	138	138

Note: This table reports the heterogeneous treatment effects on board students' personal resilience and self-esteem for different subgroups. (1)(2) for board students with day friends, (3)(4) for board students without day friends, (5)(6) for board students with friends, (7)(8) for board students without friends. Standard errors are presented in parentheses. P-values and Romano-Wolf (RW) adjusted p-values are included to account for multiple hypothesis testing. Observations vary across the models depending on the sample size for each subgroup. *p < 0.10, **p < 0.05, ***p < 0.01.

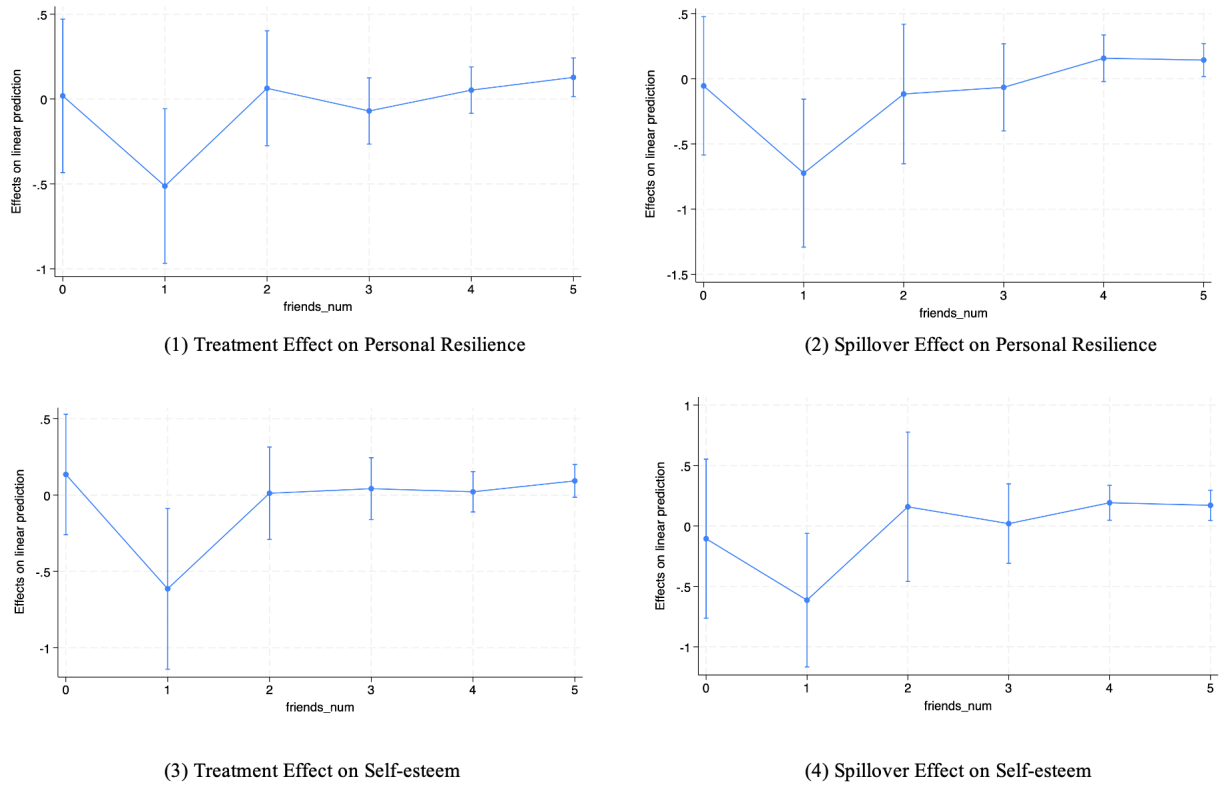
Figure 5. Heterogeneous Impact on Board Students

	(1)	(2)	(3)	(4)
	Personal Resilience	Self-esteem	Personal Resilience	Self-esteem
Treatment	0.109*	0.167***	0.085	0.078
	(0.063)	(0.055)	(0.092)	(0.092)
P value	0.090	0.004	0.361	0.404
RW p-value	0.150	0.041	0.350	0.350
Observations	2,017	2,017	877	877

Note: This table presents the treatment effects on personal resilience and self-esteem for day students, separated by those with and without board friends. The results show significant effects on personal resilience and self-esteem in columns (1) and (2), with the corresponding Romano-Wolf (RW) adjusted p-values. The sample size (observations) varies across the models. Standard errors are in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01.

Figure 6. Impact on Day Students With and Without Board Friends

The results indicate that the spillover effect is primarily driven by day students with board friends, suggesting that peer networks serve as an important channel through which the treatment influences non-treated students. Notably, only 10% of day students lack boarder friends, implying that the potential for indirect treatment effects is widespread.



(1) Treatment Effect on Personal Resilience

(2) Spillover Effect on Personal Resilience

(3) Treatment Effect on Self-esteem

(4) Spillover Effect on Self-esteem

Figure 7. Marginal Effect of Treatment on Each Friends' Number

Figure 7 illustrates the marginal effect of treatment on personal resilience and self-esteem, contingent on the number of friends. A key observation from the analysis is that the treatment effect appears larger and more significant for individuals with a greater number of friends. Research in peer effect literature suggests that individuals with more extensive social networks tend to have a larger influence on those around them (Sacerdote, 2001a; Carman and Zhang, 2012; Feld and Zölitz, 2017; Garlick, 2018; Zárate, 2023; Hu, 2023). Interestingly, our findings suggest that there is probably no need to specifically target individuals with larger networks for intervention. Instead, treating everyone equally seems to result in these individuals naturally receiving the additional treatment effect. This implies that the inherent dynamics of social networks can amplify the benefits of interventions, especially among those with a higher number of connections, reinforcing the importance of inclusive treatment strategies in fostering resilience and self-esteem.

	# Board Friends	%Female Friends	# Persistent Friends
Treatment	-0.303	-0.002	-0.179
	(0.204)	(0.005)	(0.241)
After	0.153	0.008	-2.586***
	(0.100)	(0.006)	(0.159)
Interaction	0.079	0.001	0.261
	(0.142)	(0.008)	(0.165)
P-value	0.581	0.899	0.119
RW p-value	0.664	0.891	0.149
Observations	6,297	5,962	6,297

Note: This table shows the impact of the intervention on network structure, measured by the number of board friends, the percentage of female friends, and the number of persistent friends is the number of nominated friends at baseline in 2015 and the number of those who were nominated at both baseline and follow-up survey in 2017. The results suggest no significant treatment effect on the network structure. Romano-Wolf (RW) adjusted p-values are reported to account for multiple testing, and standard errors are shown in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01.

Figure 8. No Change in Network Structure Post-Intervention

Despite evidence of peer effects, the intervention did not alter the overall social network structure within schools. As shown in Figure 8, there were no significant changes in the composition or connectivity of student networks following the intervention. The variable "After" is a time dummy, taking the value 0 at baseline and 1 in 2017. The coefficients of the interaction term capture the treatment's influence on network structures over time. The results indicate that the intervention neither affected the number of boarder friends students had nor altered the gender composition of their friendships. Furthermore, the number of "persistent friends"-those nominated at baseline who remained in the students' network in 2017-remained unchanged across both treatment and control groups, suggesting that the treatment did not impact the continuity of social connections.

This stability in friendships is consistent with the fact that students were in grade 4 and grade 5 at baseline, meaning they had known each other for 4-5 years, making a persistent friendship network a natural outcome. These findings imply that the observed spillover effects likely stem from pre-existing social ties rather than treatment-induced changes in network structure. The stability of the network allows us to interpret the peer effects as the influence of peers' non-cognitive skills on individual outcomes, rather than a conflation of changes in network formation and peer outcomes. This enhances the clarity of our interpretation and isolates the peer effect from potential confounding factors related

to network evolution.

7.2 Heterogenous Effect

Table 8. Resilience (2017)

Covariates		Treatment	Covariate	Interaction	N
Male	Board	0.126***	-0.347***	-0.044	4,699
		(0.041)	(0.040)	(0.060)	
	Day	0.065	-0.319***	0.081	2,763
		(0.052)	(0.056)	(0.078)	
Grade 4	Board	0.044	-0.214***	0.125**	4,699
		(0.041)	(0.046)	(0.060)	
	Day	0.094*	-0.212***	0.017	2,763
		(0.055)	(0.063)	(0.078)	
Left-behind	Board	0.107***	-0.014	-0.008	4,699
		(0.041)	(0.041)	(0.060)	
	Day	0.032	-0.099*	0.163**	2,763
		(0.052)	(0.058)	(0.079)	
Parents Marriage Shock	Board	0.091***	-0.181***	0.087	4,699
		(0.032)	(0.059)	(0.087)	
	Day	0.091**	0.004	0.121	
		(0.041)	(0.098)	(0.139)	2,763

Note: Robust standard errors in parentheses. Treatment is a dummy variable indicating whether randomly assigned to the audio bedtime story treatment group. Coefficients of co-variate show the initial difference in gender, first treated grade etc. in the control group. Coefficients of interaction show, the difference in treatment effects across co-variate. All regressions control for variables specified in the empirical strategy regression part. Board stands for using only board students samples and Day means the regression results for only day students samples.

Table 8 shows heterogeneous treatment effects and spillover effects on resilience in 2017 across genders, whether they were first treated in grade 4, whether they are left-behind children (at least one parent went out of their hometown to work), and whether their parents experienced a marriage shock such as divorce, death of their partner, or runaway of their partner. Results suggest that girls are more resilient after treatment. Boarding girls' resilience increased by 0.126 SDs. Boys were less resilient than girls before treatment. There is no gender difference in treatment effect and spillover effects in terms of resilience.

After 14 months of treatment, there is no significant treatment effect on board students first treated from grade 5. However, board students first treated from grade 4 got a 0.125 SD increase in resilience from treatment. The results suggest that students benefit more from being treated at an earlier stage in terms of resilience development.

Non-left-behind students with both parents' companionship benefit from treatment with a 0.107 SD increase in resilience. Results suggest there is no significant difference in treatment effects between left-behind and non-left-behind board students. The spillover effects, however, are larger among left-behind day students than non-left-behind students. The difference in spillover effects on resilience might help us understand how personality transfers among school-age children in the next chapter.

Students without experiencing a parent marriage shock realized a 0.091 SD increase in resilience after 14 months of treatment. The spillover effect on resilience is 0.091 SDs among day students with no parent marriage shock. Neither difference in treatment effects nor difference in spillover effects are significantly different between students with and without a parent marriage shock.

8 Peer Influence Evaluation

8.1 Empirical Strategy

8.1.1 Linear-in-Means Model IV Estimation

For comparison with the general model and learning how the average peer affect one's own outcome, we provide IV estimation results for the reduced-form LIM model:

$$y_i = x_i\gamma + \lambda\bar{y}_{-i} + \alpha_c + \varepsilon_i \tag{2}$$

where y_i is the outcome in noncognitive skills, and $\bar{y}_{-is}$ is the average outcome of i's nominated board friends who are directly treated ($\bar{y}_{-i} = \sum_j \tilde{g}_{ij} \cdot y_j$). Adjacency matrix $G_{n*n} = [g_{ij}]$: $g_{ij} = 1$ if i nominated j as friend otherwise 0, n_i is the number of nominated board friends, $\tilde{G}_{n*n} = [\tilde{g}_{ij}]$ where $\tilde{g}_{ij} := g_{ij}/n_i$. x_i are a vector of control variables. Under the independence assumption of the error term, control variables and network are exogenous: $\mathbb{E}[\varepsilon_i|\mathbb{Z},\mathbb{G}] = 0$. This assumption is supported by the reduced form results that the network structure does not change with the intervention.Instrumented by their characteristics x_i, treatment status T_s and their board friends' average distance from home to school d_{-i} ($d_{-i} = \sum_j \tilde{g}_{ij} \cdot d_j$), the IV estimator tells us how average peers' noncognitive skills affect an individual's noncognitive skills.

Since the audio bedtime story was delivered to boarding students through dormitory speakers, the treatment did not directly affect the outcomes of day students. Additionally, the distance between boarding students' homes and the school is unlikely to influence day students' outcomes or interactions between the two groups, as most of their peer interactions occur within the school setting. Therefore, the exclusion restrictions for the two instruments are valid. This is further supported by the overidentification test, which fails to reject the null hypothesis that the instruments are valid, as shown in the results tables.

presents the first-stage results, illustrating how the average outcomes of day students' boarder peers are influenced by the instruments: treatment and peer distance from home. The treatment increases personal resilience and self-esteem by 0.129 and 0.128 standard deviations (SD), respectively. These effects are larger than the average treatment effect

	Personal Resilience	Social Resilience	Self-esteem	Internalization	Externalization	Depression
Treatment	0.129*	0.104	0.128*	-0.003	-0.033	-0.051
	(0.071)	(0.067)	(0.074)	(0.061)	(0.081)	(0.073)
Peer Distance	0.001	0.000	0.000	-0.004	-0.003	-0.000
	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)
Cragg-Donald F-statistic	1.990	1.500	1.772	0.396	0.194	0.282
Montiel-Pflueger weak instrument robust F statistic	2.655	1.663	2.473	0.490	0.211	0.307
Observations	1,928	1,928	1,928	1,805	1,853	1,924

Note: The dependent variables are the weight average of the day student's nominated board friends' outcomes as defined earlier in the equation 5.2. Table presents OLS coefficients for the bedtime story treatment group compared to the control group, with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimates control for all control variables in the regression section of the empirical strategy. Standard errors are clustered at the school level. Cragg-Donald F-statistics assume homoskedasticity and Montiel-Pflueger F-statistics adjust for heteroscedasticity are weak instrument robust.

Figure 9. First Stage: The Effect of Treatment and Peer Distance from Home to School on Boarding Peers' Outcomes

observed for boarder students' personal resilience and self-esteem. This aligns with the observation that the treatment effect on boarders who have day student friends is greater than for those without such friendships. However, both the Cragg-Donald and Montiel-Pflueger F-statistics indicate that the instruments are weak.

To address the issue of weak instruments, we estimate the instrumental variable (IV) models using four different methods: two-stage least squares (2SLS), generalized method of moments (GMM), limited information maximum likelihood (LIML), and Fuller(1) adjusted LIML. Results are shown in Figure 10. The Fuller(1) LIML and LIML estimates are more robust to weak instruments than 2SLS and GMM. While LIML is asymptotically equivalent to 2SLS, it tends to exhibit less bias in finite samples with weak instruments. The Fuller(1) estimator, introduced by Wayne Fuller, is a modified version of LIML designed to reduce the estimator's variance, particularly in small samples and in the presence of weak instruments. Fuller(1) further improves upon LIML's performance under these conditions.

The similarity between the Fuller(1) LIML and 2SLS estimators, as well as the consistency of all four estimators in terms of coefficient values and standard errors, suggests the robustness of our findings. A one SD increase in the average peers' personal resilience and self-esteem results in an approximately 1 SD and 1.34 SD increase in an individual's personal resilience and self-esteem, respectively. Therefore, the peer effects on personal resilience and self-esteem are substantial and non-negligible.

The practical implications of these findings are significant for child development, particularly in educational and social settings. Enhancing peer interactions among students from different backgrounds can foster important personal development traits, such as resilience and self-esteem. Policies that encourage structured peer engagement and mixed social environments, such as integrating day and boarder students, may have a profound impact on children's emotional and psychological growth. Educational policymakers should consider interventions that promote social inclusion and

peer support programs as a means to bolster these traits, which are critical for long-term success both academically and personally. Additionally, understanding the role of peer effects in shaping these developmental outcomes can inform the design of school curricula and extracurricular activities that intentionally create opportunities for positive peer influence.

	(1)	(2)	(3)	(4)	(9)	(10)	(11)	(12)
	Personal Resilience				Self-esteem			
	2SLS	GMM	LIML	Fuller	2SLS	GMM	LIML	Fuller
Average Peer Influence	0.975**	0.893*	1.062*	0.995**	1.345	1.386**	1.441**	1.343**
	(0.489)	(0.466)	(0.573)	(0.508)	(0.643)	(0.661)	(0.741)	(0.641)
Overidentifying test statistic	1.393	1.447	1.352	1.080	1.026	0.921	0.990	0.704
Overidentifying test P-value	0.248	0.229	0.245	0.299	0.311	0.337	0.320	0.401
Observations	1,928	1,928	1,928	1,928	1,928	1,928	1,928	1,928

Note: the dependent variables are personal resilience and self-esteem factor scores. Average peer influence is the weight average of the day student's nominated board friends personal resilience and self-esteem respectively. Table presents the estimators by two-stage-least-square, GMM, limited information maximum likelihood and fuller(1) adjusted LIML estimation where the last two estimations are weak instrument robust , with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimates control for all control variables in the regression section of the empirical strategy. Standard errors are clustered at the school level.Overidentifying tests are Sargan test for 2SLS, Hansen test for GMM and Anderson-Rubin Test for the last two LIML estimation. All overidentifying tests suggest we cannot reject valid instruments.

Figure 10. IV Estimation: Effect of Boarding Peers' Outcomes on Day Students' Outcomes

9 Conclusions and Direction for Future Research

This paper has provided comprehensive evidence on both the direct and spillover effects of delivering audio bedtime stories to rural boarding school students in China, with a focus on non-cognitive skill development, particularly resilience and self-esteem. The results reveal a positive impact on both treated students and their untreated peers, emphasizing the pivotal role of friendship networks in facilitating these spillover effects. Students with more treated friends, or a larger number of friends overall, experienced greater gains in personal resilience and self-esteem. A one standard deviation increase in the average personal resilience or self-esteem of nominated friends resulted in an approximate increase of 0.9-1 SD in personal resilience and 1.35-1.45 SD in self-esteem for an individual. These findings underscore the significant influence of peer networks on the development of non-cognitive skills, broadening our understanding of how social connections contribute to personal growth in educational settings.

The intent-to-treat (ITT) analysis shows that while the intervention initially led to a short-term increase in depression, particularly among boys, this effect diminished over time. Furthermore, no significant spillover effects on depression were observed among untreated peers. In contrast, the positive spillover effects on resilience and self-esteem among untreated day students were larger than the direct treatment effects observed in treated boarding students. This can

partly be explained by the fact that board students with day-student friends experienced greater treatment effects than those without day-student friends, meaning the treatment effect on day students' friends exceeded the average treatment effect. Additionally, the higher baseline levels of resilience and self-esteem in day students may have made them more responsive to peer influences.

These findings align with the broader literature on peer effects and non-cognitive skill development, particularly within educational contexts. This study contributes new insights by showing that peer outcomes,Äîrather than simply peer characteristics-can significantly influence untreated students, offering empirical evidence that non-cognitive skills are transferable through social networks. This is an important finding, suggesting that educational interventions targeting a select group of students can produce wider benefits through peer-to-peer interactions, amplifying the intervention's overall impact.

10 Future Research Directions

Future research should explore non-linear models to better understand how peer influence operates in different contexts. The linear-in-means model used in this study could be supplemented by other approaches, such as the "one bad apple" model (where the most disruptive peer has the strongest influence) or the "shining light" model (where the highest-performing peer drives outcomes).

We are interested in the peer influence weight and how peers with different performances affect an individual differently. The production function can be expressed as:

$$y_i = x_i\gamma + \lambda\tilde{y}_{-i} + \alpha_c + \varepsilon_i \tag{3}$$

where $\tilde{y}_{-i}$ is a function of individual i's peer outcomes. λ is the peer-effect propagation rate, α_c is the county fixed effect. By modelling peer effects within a general framework in the human capital production function, the analysis can evaluate whether the linear-in-means, shining light, or bad apple model best explains peer influences on noncognitive skills. Identifying which model best captures peer influence dynamics will have important implications for educational policy and practice. For instance, if the "shining light" model is more effective, schools might focus on fostering networks around top-performing students. Conversely, if the "bad apple" model proves more relevant, interventions could target lower-performing students to minimize their negative impact on peers.

Additionally, some studies were limited by the lack of detailed network data, which restricted the definition of peers to classmates or cohort members(Feng et al., 2024; Cattan et al., 2023). Future research should compare the peer effects estimated with and without network data would provide valuable insights into how the absence of such data may bias

our understanding of peer dynamics and educational outcomes.

In conclusion, the findings of this study highlight the potential for low-cost, scalable interventions to significantly enhance students' non-cognitive development, not only among those directly treated but also through their peer networks. As policymakers and educators work to address challenges in non-cognitive skill development, such as resilience and self-esteem, interventions that leverage peer influence offer a promising strategy for improving student outcomes, particularly in under-resourced settings.

11 Appendices

In this section, we show the full-sample balance test, the Lee Bounds Estimates of ITT effect for board students in 2017, and the balance test for board students in 2017.

11.1 Survey Scales

In this section, we first describe the other non-cognitive measures used in the study, followed by a presentation of the survey questions for each measure.

The internalizing and externalizing behavior scales were adapted from the "Gansu Basic Education Tracking Survey Project." The internalizing scale consists of 18 items, while the externalizing scale includes 17 items, both of which have been validated for use among Chinese primary school students.

The self-esteem scale is based on the Rosenberg (1965) Self-Esteem Scale, which consists of 10 items. Higher scores reflect higher self-esteem, with scores above 15 considered to be within the normal range.

The depression risk scale was adapted from the CES-DC scale, developed by the American Children's Depression Research Center based on Radloff (1977). This scale includes 20 items and measures children's emotional experiences over the past week. Higher scores indicate a greater risk of depression, with scores above 15 denoting "depression risk" and scores above 30 indicating "high depression risk".

Table 9. Full Sample Balance Test

	Treat	Control	$H_0 : 1 = 2$
Individual Characteristics:			
Age	10.30	10.35	0.59
	[0.85]	[0.83]	
Gender	0.50	0.53	0.13
	[0.50]	[0.50]	
Grade	0.49	0.48	0.74
	[0.50]	[0.50]	
Weight	32.56	33.46	0.28
	[7.51]	[8.32]	
Height	138.17	138.88	0.44
	[7.73]	[7.63]	
Reading Score	79.23	80.32	0.23
	[7.45]	[8.04]	
Family Characteristics:			
Mother's Education	8.82	8.71	0.82
	[2.69]	[2.73]	
Father's Education	9.03	8.96	0.66
	[2.53]	[2.55]	
Mother's Health	1.68	1.64	0.72
	[0.79]	[0.79]	
Father's Health	1.63	1.58	0.46
	[0.79]	[0.79]	
Parents' Migration	1.00	0.89	0.69
	[1.27]	[1.21]	
Parents' Marriage	1.25	1.26	0.19
	[0.77]	[0.77]	
Sibling Count	2.18	2.20	0.4
	[0.90]	[0.90]	
Family Assets	0.02	0.16	0.69
	[2.33]	[2.29]	
Outcome Variables:			
Resilience	128.30	128.52	0.51
	[15.27]	[14.70]	
Depression	18.92	19.09	0.39
	[8.99]	[8.84]	
Sleeping Score	35.55	35.41	0.32
	[5.20]	[5.03]	

Note: Table shows balance test for board students and day students pooled together. The outcome variables are Measurement Scores(not standardized).

Table 10. Lee Bounds Estimates of ITT Effects in 2017 for Board Students

	Lower	Upper	Lower 95% CI	Upper 95% CI
Resilience	-0.016	-0.004	-0.074	0.058
	(0.030)	(0.032)		
Optimism	-0.007	0.004	-0.067	0.064
	(0.030)	(0.030)		
Self-efficacy	-0.008	0.004	-0.062	0.057
	(0.028)	(0.027)		
Adult relationships	-0.006	0.001	-0.066	0.061
	(0.031)	(0.030)		
Peer relationships	-0.003	0.004	-0.059	0.062
	(0.029)	(0.030)		
Interpersonal sensitivity	-0.008	0.002	-0.061	0.057
	(0.027)	(0.028)		
Emotional regulation	-0.005	0.006	-0.058	0.058
	(0.027)	(0.027)		
Internalization	-0.011	0.002	-0.069	0.063
	(0.030)	(0.031)		
Externalization	-0.011	0.005	-0.063	0.059
	(0.027)	(0.028)		
Self-esteem	-0.006	0.005	-0.071	0.068
	(0.033)	(0.032)		

Note: Table shows Lee's bound results for specified characteristics and baseline outcomes for board students in 2017. Control variables are specified in the empirical strategy part.

Table 11. Balance Test for Board Students in 2017

	Board Student		
	Control	Treat	P-value/ RW p-value
Personal Characteristics:			
Age	11.956	11.922	0.167
	(0.813)	(0.846)	[0.821]
Gender	0.520	0.505	0.310
	(0.500)	(0.500)	[0.956]
Grade	5.539	5.516	0.123
	(0.499)	(0.500)	[0.742]
Height	147.596	146.967	0.012
	(8.582)	(8.712)	[0.144]
Weight	38.899	38.733	0.525
	(8.971)	(8.954)	[0.992]
Family Characteristics:			
Mother's Education	5	8.223	0.509
	(2.166)	(2.170)	[0.992]
Father's Education	8.680	8.655	0.686
	(2.102)	(2.094)	[0.999]
Mother's Health	1.611	1.665	0.020
	(0.777)	(0.818)	[0.210]
Father's Health	1.556	1.610	0.019
	(0.771)	(0.806)	[0.210]
Parents' Migration	0.479	0.477	0.905
	(0.500)	(0.500)	[0.999]
Parents' Marriage	0.140	0.144	0.683
	(0.347)	(0.352)	[0.999]
Baseline Outcome Variables:			
Baseline Resilience	127.451	128.156	0.320
	(15.120)	(14.412)	[0.956]
Baseline Depression	19.539	19.710	0.687
	(8.904)	(8.901)	[0.999]
Baseline Sleeping score	35.810	35.889	0.788
	(5.109)	(4.922)	[0.999]

Note: Table shows balance test for specified characteristics and baseline outcomes for board students in 2017. Outcome variables are measurement scores(not standardized).

Table 12. Six Dimensions of Resilience

Direction	Characteristics	Question
Positive	Adult relationships	1. At least one adult can listen to me about my problem
Positive	Peer relationships	2. I make friends easily
Negative	Adult relationships	3. I don't believe adults
Positive	Self-efficacy	4. I am proud to stand up for my beliefs
Positive	Interpersonal sensitivity	5. If I don't like other people, I will try to express it tactfully so that him does not get hurt
Negative	optimism	6. I feel bad more than I feel good
Positive	Interpersonal sensitivity	7. I apologize when I accidentally hurt or offend someone
Positive	optimism	8. I will encounter more good things than bad things
Positive	Self-efficacy	9. When there are many things to think about and do, I will come little by little, one by one.
Negative	Peer relationships	10. I don't like to be with friends
Positive	Adult relationships	11. There are some adults I respect and admire
Positive	Self-efficacy	12. I try to see and understand the problem from different angles
Negative	Interpersonal sensitivity	13. I don't like helping others
Negative	Emotional control	14. Even small things frustrate me
Positive	Peer relationships	15. I find it interesting to be with friends Completely
Negative	Adult relationships	16. Adults usually ignore me Fully
Negative	optimism	17. Most of the time I come together and think that this day will not go well
Positive	Self-efficacy	18. If the way to deal with things is not right, I will try to think of other ways
Positive	Self-efficacy	19. I am as important as everyone else
Negative	Peer relationships	20. I don't have a trusted friend

Students answer a number from 1 to 4. 1=Strongly disagree, 2=Disagree, 3=Agree, and 4=Agree wholeheartedly.

Table 13. Six Dimensions of Resilience (Cont.)

Direction	Characteristics	Question
Positive	optimism	21. Even if there are bad things, I can still see the good side of myself and my life
Negative	optimism	22. I think most things in life are boring
Negative	Emotional control	23. I keep making the same mistake
Negative	Adult relationships	24. I don't have much respect for authority figures
Positive	Self-efficacy	25. I am satisfied with the choices I have made in my life
Negative	optimism	26. I think my future will get worse and worse
Positive	Self-efficacy	27. I do a lot of things well
Negative	Emotional control	28. I don't have the patience to do things
Negative	Emotional control	29. I make a decision without thinking about the consequences
Positive	Emotional control	30. I can stay calm even if the situation is bad
Negative		31. I stick to what I want to do and don't care about anything else
Positive	Adult relationships	32. In addition to my parents, there are other adults who have influenced me
Positive	Optimism	33. I consider myself a lucky person
Positive	Interpersonal sensitivity	34. Others think I understand them well Fully
Negative	optimism	35. When something bad happens, I think it will be there for a while
Negative	Peer relationships	36. I don't have many friends
Positive	Self-efficacy	37. I will do well on school exams
Positive	Adult relationships	38. If I encounter a problem, I will discuss it with an adult
Negative	Peer relationships	39. I am not welcome among friends
Negative	Peer relationships	40. I am not welcomed among my friends
Positive	Adult relationships	41. I listen to adults
Positive	Interpersonal sensitivity	42. I am good at listening to others
Positive	Peer relationships	43. I have a good relationship with my friends
Positive	Interpersonal sensitivity	44. I try to speak from others' perspectives

Students answer a number from 1 to 4. 1=Strongly disagree, 2=Disagree, 3=Agree, and 4=Agree wholeheartedly.

Table 14. Depression Questions

Direction	Question
Positive	1. Things that don't usually bother me make me feel annoyed
Positive	2. I don't want to eat and I don't have a good appetite
Positive	3. I don't feel happy, even if family and friends want me to feel better
Negative	4. I feel like I'm as good as any other kid
Positive	5. I feel like I can't concentrate on what I'm doing
Positive	6. I feel depressed and unhappy
Positive	7. I feel too tired to do things
Negative	8. I feel like something good is going to happen
Positive	9. I feel like I've not done anything in the past as I wanted
Positive	10. I feel scared
Positive	11. I don't sleep well
Negative	12. I'm happy
Positive	13. I'm quieter than usual
Positive	14. I feel lonely, I don't seem to have many friends
Positive	15. The kids I know are not friendly, they don't play with me
Negative	16. I'm happy
Positive	17. I feel like I want to cry
Positive	18. I feel sad and sad
Positive	19. I think people don't like me
Positive	20. I have difficulty making progress on anything

Students answer a number from 1 to 4. 1 = no at all (0 times), 2=Occasionally (1-2 times), 3=Sometimes (3-4 times), and 4 = Often (5-7 times).

Table 15. BFI Questions

Direction	Characteristics	Question
Positive	Extroversion	1. I'm a talkative person myself
Negative	Easy-going	2. I am a person who is critical of others
Positive	Responsibility	3. I am a person who can do all kinds of jobs
Positive	Nervousness	4. I am a depressed, depressed and sad person
Positive	Open	5. I'm creative and can come up with new ideas
Negative	Extroversion	6. I am a conservative and introverted person
Positive	Easy-going	7. I am a helpful and selfless person
Negative	Responsibility	8. I'm a bit careless
Negative	Nervousness	9. I am relaxed and can handle stress well
Positive	Open	10. I'm curious about different things
Positive	Extroversion	11. I am an energetic person
Negative	Easy-going	12. I tend to provoke conflicts
Positive	Responsibility	13. I am a reliable worker
Positive	Nervousness	14. I am a nervous person
Positive	Open	15. I am a smart, deep thinker
Positive	Extroversion	16. I am a passionate person
Positive	Easy-going	17. I am a person with a tolerant nature
Negative	Responsibility	18. I am a disorganized person
Positive	Nervousness	19. I am a person who is worried about many things
Positive	Open	20. I am imaginative

Students answer a number from 1 to 5. 1=strongly disagree, 2=somewhat disagree, 3=neither agree nor disagree, 4=somewhat agree, and 5=strongly agree.

Table 16. BFI Questions (Cont.)

Direction	Characteristics	Question
Negative	Extroversion	21. I am a person who tends to be quiet
Positive	Easy-going	22. I am someone who easily trusts others
Negative	Responsibility	23. I'm a rather lazy person
Negative	Nervousness	24. I am emotionally stable and not easily upset
Positive	Open	25. I am a creative person
Positive	Extroversion	26. I am an assertive person
Negative	Easy-going	27. I am a ruthless person
Positive	Responsibility	28. I am the one who can persevere until the task is completed
Positive	Nervousness	29. I am moody
Positive	Open	30. I am someone who values artistic and aesthetic experience
Negative	Extroversion	31. I am sometimes shy and restrained
Positive	Easy-going	32. I am a thoughtful and friendly person to almost everyone
Positive	Responsibility	33. I am an efficient person
Negative	Nervousness	34. I am the one who keeps calm in a stressful situation
Negative	Open	35. I am someone who likes to work step by step
Positive	Extroversion	36. I am outgoing and sociable
Negative	Easy-going	37. I am sometimes rude to others
Positive	Responsibility	38. I am the one who makes the plan and follows through it
Positive	Nervousness	39. I am a person who is easily anxious and nervous
Positive	Open	40. I am someone who likes to reflect and have an active mind
Negative	Open	41. I am someone who has little interest in art
Positive	Easy-going	42. I am someone who likes to work with others
Negative	Responsibility	43. I am easily distracted
Positive	Open	44. I am a person who is well-versed in art, music, and literature

Students answer a number from 1 to 5. 1=strongly disagree, 2=somewhat disagree, 3=neither agree nor disagree, 4=somewhat agree, and 5=strongly agree.

Table 17. School Environment Questions

Direction	Question
Positive	1. Most students turn in their assignments on time
Positive	2. Teachers treat students from different backgrounds with the same respect
Positive	3. School rules are reasonable
Positive	4. This school campus is safe
Positive	5. The school rules of this school are clearly communicated to students
Positive	6. Most students try their best to achieve the best
Positive	7. Teachers care about students
Positive	8. The school's penalties for violations of the rules and regulations are reasonable
Negative	9. There are students threatening and bullying other students
Positive	10. Students know what they should do in school
Positive	11. Students are friendly to each other
Positive	12. Most students feel happy
Positive	13. Students feel safe
Negative	14. Students are worried that others are bullying them
Positive	15. Students understand the rules and regulations of the school
Positive	16. Students care about each other
Positive	17. When students are in difficulty, the teacher listens to them
Positive	18. School rules and regulations are reasonable
Positive	19. Students know they are safe in this school
Positive	20. Students know how to behave in school

Students answer a number from 1 to 4. 1=strongly disagree, 2=disagree, 3=agree, and 4=strongly agree.

Table 18. School Environment Questions (Cont.)

Direction	Question
Positive	21. Mutual respect between students from different backgrounds
Positive	22. Adults working at schools care about students
Positive	23. Most students follow the school rules
Negative	24. Bullying is a big problem in this school
Positive	25. Most students like this school
Positive	26. Teachers like their students
Negative	27. Students bullying each other
Positive	28. Class rules are fair
Positive	29. Most students study hard to get good grades
Positive	30. Mutual respect among students
Positive	31. Students get along well
Positive	32. In this questionnaire, I expressed my true thoughts

Students answer a number from 1 to 4. 1=strongly disagree, 2=disagree, 3=agree, and 4=strongly agree.

Table 19. Bullying Questions

Direction	Question
Positive	A. In the past six months, I have seen others being bullied in school
Positive	B. In the past six months, the number of times I have been bullied in school
Positive	C. In the past six months, the number of times I have bullied others in school
Positive	1. People used to call me an ugly nickname
Positive	2. Others have criticized me on the Internet
Positive	3. Others have scolded me with swear words
Positive	4. Someone else has destroyed my belongings 1
Positive	5. Others have united with peers to exclude me
Positive	6. Others have said bad things about me in public
Positive	7. Someone has threatened me with words
Positive	8. Others have forced me to do things
Positive	9. Others have spread rumours about me on the Internet
Positive	10. Others used to laugh at me
Positive	11. Others once made small reports to the teacher to embarrass me
Positive	12. Others have posted indecent photos of me on the Internet
Positive	13. Others have deliberately pushed me or tripped me
Positive	14. Others have taken my belongings in private

Students answer a number from 1 to 5. 1 = not at all, 2 = 1-2 times in half a year, 3 = 2-3 times a month, 4 = 1 time per week, and 5 = several times a week. If the student's average score is greater than 2, it means that the student has witnessed a bullying incident, experienced a bullying incident, or bullied others.

Table 20. Questions about Reading Habits and Interest

Question
1. How many books do you have at home (A=0-10 B=11-25 C=26-50 D=51-100 E=101-200 F=201-500 G=500 or more)
2. How long do you read extracurricular books out of interest and hobby every day from Monday to Friday (A = I don't read extracurricular books B = less than 30 minutes C = 30 minutes - 1 hour D = 1-2 hours E = more than 2 hours)
3. How long will you read extracurricular books out of interest and hobby every day on Saturday and Sunday (A = I don't read extracurricular books B = less than 30 minutes C = 30 minutes - 1 hour D = 1-2 hours E = more than 2 hours)
4. I only read books as a last resort
5. Reading is one of my favorite things
6. I like to discuss books with others
7. I find it difficult to finish a book
8. I am very happy when I receive a book as a gift
9. Reading is simply a waste of time for me
10. Going to a bookstore or library is a good pleasure
11. I read only to get the necessary information
12. I can't calm down and read for no more than a few minutes
13. I like to say what I think about books I've already read
14. I like to exchange books with my friends

If not otherwise noted, students answer a number from 1 to 4. 1=Disagree Strongly, 2=Disagree, 3=Agree, and 4=Agree wholeheartedly.

11.2 Robustness Checks

In this section, we show the average treatment and spillover effect using overall scores. The result is consistent with our findings in the main body.

	Control Mean	Treatment	P-value RW p-value	Observations
		Panel A: 2016		
Resilience	0.001	0.030	0.565	3,705
		(0.051)	[0.981]	
Optimism	0.041	-0.031	0.488	5,056
		(0.044)	[0.990]	
Self-efficacy	-0.107	-0.044	0.221	5,067
		(0.036)	[0.986]	
Adult Relationships	0.006	0.060	0.182	5,056
		(0.044)	[0.643]	
Peer Relationships	0.007	-0.024	0.584	5,056
		(0.043)	[0.997]	
Interpersonal Sensitivity	-0.131	0.037	0.429	5,060
		(0.047)	[0.929]	
Emotional Regulation	0.107	0.031	0.411	5,057
		(0.037)	[0.931]	
Internalization	-0.158	-0.039	0.421	4,420
		(0.048)	[0.967]	
Externalization	-0.158	-0.027	0.503	4,488
		(0.040)	[0.997]	
Self-esteem	0.099	0.010	0.839	4,746
		(0.049)	[0.997]	
		Panel B: 2017		
Resilience	0.056	0.073*	0.099	4,337
		(0.048)	[0.700]	
Optimism	0.033	0.041	0.154	5,085
		(0.029)	[0.969]	
Self-efficacy	-0.049	-0.066**	0.017	5,097
		(0.027)	[0.764]	
Adult Relationships	0.050	0.057*	0.051	5,085
		(0.029)	[0.878]	
Peer Relationships	0.062	0.025	0.374	5.086
		(0.028)	[0.990]	
Interpersonal Sensitivity	0.066	0.029	0.318	5,085
		(0.029)	[0.997]	
Emotional Regulation	0.052	0.075***	0.007	5,090
		(0.028)	[0.641]	
Internalization	-0.112	-0.013	0.653	4,812
		(0.029)	[0.997]	
Externalization	-0.126	-0.044	0.102	4,836
		(0.027)	[0.990]	
Self-esteem	0.126	0.034	0.267	4,906
		(0.031)	[0.990]	

Note: Table presents OLS coefficients for the bedtime story treatment group compared to the control group, with standard errors in parentheses. Dependent variables are standardized measurement scores. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. RW P values use the Romano-Wolf method to account for multiple hypothesis testing for all outcome measures in this chapter. Estimates control for all control variables in the regression section of the empirical strategy.

Figure 11. Treatment effect on board students noncognitive skills (overall score)

	Control Mean	Treatment	P-value RW p-value	Observations
		Panel A: 2016		
Resilience	0.100	0.064*	0.100	3,123
		(0.044)	[0.632]	
Optimism	0.151	0.056	0.113	3,123
		(0.035)	[0.593]	
Self-efficacy	-0.122	-0.049	0.123	3,128
		(0.031)	[0.666]	
Adult Relationships	0.063	0.022	0.601	3,123
		(0.041)	[0.973]	
Peer Relationships	0.085	0.050	0.183	3,123
		(0.037)	[0.616]	
Interpersonal Sensitivity	-0.020	0.000	0.995	3,123
		(0.051)	[0.980]	
Emotional Regulation	0.160	0.088**	0.035	3,123
		(0.041)	[0.343]	
Internalization	-0.259	-0.069	0.273	2,729
		(0.062)	[0.365]	
Externalization	-0.246	-0.067	0.177	2,795
		(0.049)	[0.354]	
Self-esteem	0.198	0.058	0.152	2,940
		(0.040)	[0.662]	
		Panel B: 2017		
Resilience	0.113	0.095*	0.054	2,894
		(0.048)	[0.235]	
Optimism	0.116	0.051	0.270	2,894
		(0.045)	[0.740]	
Self-efficacy	-0.105	-0.021	0.609	2,902
		(0.042)	[0.924]	
Adult Relationships	0.063	0.078	0.186	2,894
		(0.059)	[0.354]	
Peer Relationships	0.125	0.052	0.339	2,894
		(0.054)	[0.740]	
Interpersonal Sensitivity	0.079	0.101**	0.033	2,894
		(0.046)	[0.354]	
Emotional Regulation	0.112	0.013	0.700	2,896
		(0.033)	[0.924]	
Internalization	-0.189	0.021	0.687	2,710
		(0.052)	[0.980]	
Externalization	-0.243	0.049	0.235	2,894
		(0.041)	[0.924]	
Self-esteem	0.195	0.110*	0.054	2,894
		(0.056)	[0.250]	

Note: Table presents OLS coefficients for the bedtime story treatment group compared to the control group, with standard errors in parentheses. Dependent variables are standardized measurement scores. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. RW P values use the Romano-Wolf method to account for multiple hypothesis testing for all outcome measures in this chapter. Estimates control for all control variables in the regression section of the empirical strategy.

Figure 12. Spillover effect on board students noncognitive skills (overall score)

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.

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics 134*(3), 1121–1162.

Angrist, J. D. and K. Lang (2004). Does school integration generate peer effects? evidence from boston's metco program. *American Economic Review 94*(5), 1613–1634.

Attanasio, O., Á. de Paula, and A. Toppeta (2024). Intergenerational mobility in socio-emotional skills. *Working Paper*.

Attanasio, O., Á. De Paula, and A. Toppeta (2024). The persistence of socio-emotional skills life cycle and intergenerational evidence.

Balart, P., M. Oosterveen, and D. Webbink (2018). Test scores, noncognitive skills and economic growth. *Economics of Education Review 63*, 134–153.

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2015). The miracle of microfinance? evidence from a randomized evaluation. *American economic journal: Applied economics 7*(1), 22–53.

Banerjee, A. V. and E. Duflo (2007). The economic lives of the poor. *Journal of economic perspectives 21*(1), 141–167.

Barrera-Osorio, F. and L. L. Linden (2009). The use and misuse of computers in education: evidence from a randomized experiment in colombia. *World Bank Policy Research Working Paper* (4836).

Baumert, A., M. Schmitt, M. Perugini, W. Johnson, G. Blum, P. Borkenau, G. Costantini, J. J. Denissen, W. Fleeson, B. Grafton, et al. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality 31*(5), 503–528.

Bietenbeck, J. (2021). Peer motivation, and educational success. Technical report, Working Paper.

Blake, J. and N. Maiese (2008). No fairytale... the benefits of the bedtime story. *The psychologist*.

Booij, A. S., E. Leuven, and H. Oosterbeek (2017). Ability peer effects in university: Evidence from a randomized experiment. *The review of economic studies 84*(2), 547–578.

Borghuis, J., J. J. Denissen, D. Oberski, K. Sijtsma, W. H. Meeus, S. Branje, H. M. Koot, and W. Bleidorn (2017). Big five personality stability, change, and codevelopment across adolescence and early adulthood. *Journal of personality and social psychology 113*(4), 641.

Boucher, V., M. Rendall, P. Ushchev, and Y. Zenou (2024). Toward a general theory of peer effects. *Econometrica 92*(2), 543–565.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of econometrics 150*(1), 41–55.

Buser, T., M. Niederle, and H. Oosterbeek (2021). Can competitiveness predict education and labor market outcomes? evidence from incentivized choice and survey measures. Technical report, National Bureau of Economic Research.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The review of economic studies 76*(4), 1239–1267.

Carman, K. G. and L. Zhang (2012). Classroom peer effects and academic achievement: Evidence from a chinese middle school. *China Economic Review 23*(2), 223–237.

Carrell, S. E. and M. L. Hoekstra (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics 2*(1), 211–228.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Cattan, S., K. G. Salvanes, and E. Tominey (2023). First generation elite: the role of school networks. Technical report, IFS Working Papers.

De Giorgi, G. and M. Pellizzari (2014). Understanding social interactions: Evidence from the classroom. *The Economic Journal 124*(579), 917–953.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American economic review 101*(5), 1739–1774.

Evans, W. N., W. E. Oates, and R. M. Schwab (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy 100*(5), 966–991.

Feld, J. and U. Zölitz (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics 35*(2), 387–428.

Feng, S., J. H. Kim, and Z. Yang (2024). Effects of childhood peers on personality traits.

Figlio, D. N. (2007). Boys named sue: Disruptive children and their peers. *Education finance and policy 2*(4), 376–394.

Franke, R. H. and J. D. Kaul (1978). The hawthorne experiments: First statistical interpretation. *American sociological review*, 623–643.

Fryer, R. G. and L. F. Katz (2013, May). Achieving Escape Velocity: Neighborhood and School Interventions to Reduce Persistent Inequality. *American Economic Review 103*(3), 232–237.

Garlick, R. (2018). Academic peer effects with different group assignment policies: Residential tracking versus random assignment. *American Economic Journal: Applied Economics 10*(3), 345–369.

Glewwe, P., E. Hanushek, S. Humpage, and R. Ravina (2011, October). School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010. Technical Report w17554, National Bureau of Economic Research, Cambridge, MA.

Glewwe, P., Q. Huang, and A. Park (2017). Cognitive skills, noncognitive skills, and school-to-work transitions in rural china. *Journal of Economic Behavior & Organization 134*, 141–164.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Gordanier, J., W. Hauk, and C. Sankaran (2019, October). Early intervention in college classes and improved student outcomes. *Economics of Education Review 72*, 23–29.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour economics 19*(4), 451–464.

Hennecke, M., W. Bleidorn, J. J. Denissen, and D. Wood (2014). A three–part framework for self–regulated personality development across adulthood. *European Journal of Personality 28*(3), 289–299.

Hoeve, M., G. J. J. Stams, C. E. Van der Put, J. S. Dubas, P. H. Van der Laan, and J. R. Gerris (2012). A meta-analysis of attachment to parents and delinquency. *Journal of abnormal child psychology 40*, 771–785.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika 30*, 179–185.

Hou, H., Y. Wu, Y. Song, et al. (2018). Low-age boarding and rural pupils' human capital accumulation: evidence from the" primary school merger program". *China Rural Economy* (7).

Hoxby, C. M. (2000). Peer effects in the classroom: Learning from gender and race variation.

Hoxby, C. M. and G. Weingarth (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Technical report, Citeseer.

Hsieh, C.-S. and L. F. Lee (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics 31*(2), 301–319.

Hu, Q. (2023). Breaking down bullying: Empathy, social networks, and adolescents.

Hutton, J. S., T. Horowitz-Kraus, A. L. Mendelsohn, T. DeWitt, S. K. Holland, and C.-M. A. Consortium (2015). Home reading environment and brain activation in preschool children listening to stories. *Pediatrics 136*(3), 466–478.

Jingzhong, Y. and P. Lu (2011). Differentiated childhoods: impacts of rural labor migration on left-behind children in china. *The Journal of peasant studies 38*(2), 355–377.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement 20*(1), 141–151.

Kautz, T., J. J. Heckman, R. Diris, B. Ter Weel, and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success.

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.

Lavy, V. and A. Schlosser (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics 3*(2), 1–33.

Leaver, C., O. Ozier, P. Serneels, and A. Zeitlin (2021). Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from rwandan primary schools. *American economic review 111*(7), 2213–2246.

List, J. A., F. Momeni, and Y. Zenou (2020). The social side of early human capital formation: Using a field experiment to estimate the causal impact of neighborhoods. Technical report, National Bureau of Economic Research.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: Experimental evidence from india. *Journal of political Economy 119*(1), 39–77.

Nakano, Y. and E. F. Magezi (2020). The impact of microcredit on agricultural technology adoption and productivity: Evidence from randomized control trial in tanzania. *World Development 133*, 104997.

Oosterbeek, H. and R. Van Ewijk (2014). Gender peer effects in university: Evidence from a randomized experiment. *Economics of Education Review 38*, 51–63.

Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement 1*(3), 385–401.

Robins, R. W., J. L. Tracy, K. Trzesniewski, J. Potter, and S. D. Gosling (2001). Personality correlates of self-esteem. *Journal of research in personality 35*(4), 463–482.

Romano, J. P. and M. Wolf (2005a). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*(469), 94–108.

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

Romano, J. P. and M. Wolf (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters 113*, 38–40.

Rosenberg, M. (1965). Rosenberg self-esteem scale (rse). *Acceptance and commitment therapy. Measures package 61*(52), 18.

Sacerdote, B. (2001a, May). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics 116*(2), 681–704. Publisher: Oxford University Press / USA.

Sacerdote, B. (2001b). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics 116*(2), 681–704.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annu. Rev. Econ. 6*(1), 253–272.

Shan, X. and U. Zölitz (2022). Peers affect personality development.

Song, M. (2003). *Two studies on the Resilience Inventory (RI): Toward the goal of creating a culturally sensitive measure of adolescence resilience*. Harvard University.

Song, Y. and D. Zhongchen (2018). Analysis on the influencing factors of decision-making decision of rural left-behind middle school students and their policy implications [j]. *Educational Science Research 9*, 49–55.

Spybrook, J., H. Bloom, R. Congdon, C. Hill, A. Martinez, S. Raudenbush, and A. TO (2011). Optimal design plus empirical evidence: Documentation for the ,Äúoptimal design,Äù software. *William T. Grant Foundation. Retrieved on November 5*, 2012.

Stieger, M., C. Flückiger, D. Rüegger, T. Kowatsch, B. W. Roberts, and M. Allemand (2021). Changing personality traits with the help of a digital personality change intervention. *Proceedings of the National Academy of Sciences 118*(8), e2017548118.

Velicer, W. F. and D. N. Jackson (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate behavioral research 25*(1), 1–28.

Wang, S., X. Dong, and Y. Mao (2017). The impact of boarding on campus on the social-emotional competence of left-behind children in rural western china. *Asia Pacific Education Review 18*, 413–423.

Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review 95*(2), 199–203.

Wu, H. and R. Estabrook (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika 81*(4), 1014–1045.

Zárate, R. A. (2023). Uncovering peer effects in social and academic skills. *American Economic Journal: Applied Economics 15*(3), 35–79.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics 85*(1), 9–23.