Unintended Consequences of Best Intentions: Examining Spillover Effects in Targeted Supplementary Education Interventions

Yujuan Gao, Yue Ma, Conner Mullally, and Scott Rozelle*

December 23, 2025

Frequently updated. Please click here for the latest draft.

Abstract

This study examines spillover effects of targeted educational interventions through a field experiment in 130 rural Chinese boarding schools, comparing computer-assisted learning (CAL) and traditional workbooks. Results reveal significant negative spillovers of workbook interventions on nontarget students' performance, particularly affecting those closely connected to targeted students. Effects intensify with increased exposure and peer interaction. The key mechanism appears motivational: Observing peers receiving supplementary workbook resources in class reduces students' confidence in the value of their academic efforts for future careers. CAL interventions, conducted outside classrooms, show no such spillovers, highlighting the importance of considering unintended consequences in competitive, resource-limited environments.

Keywords: Supplementary Intervention, Spillover Effect, Field Experiment

JEL Codes: I21, C93, D62

*Yujuan Gao is an PhD candidate of Food and Resource Economics at University of Florida (yujuan.gao@ufl.edu). Yue Ma is a PhD student and research fellow at Stanford University. Conner Mullally is an Associate Professor of Food and Resource Economics at University of Florida. Scott Rozelle is a Senior Fellow at Stanford University. We thank the participants and organizers of the 2025 ASSA Meeting and the 2024 Stanford Center on China's Economy and Institutions (SCCEI) Young Researcher Workshop. This randomized controlled trial is registered in the American Economic Association's registry for randomized controlled trials under ID AEARCTR-0003086.

1 Introduction

The design of economic policy interventions under resource constraints inherently involves trade-offs between efficiency and equity. To maximize effect with limited resources, interventions typically target the populations that would benefit the most from them. Nevertheless, social interactions between targeted and nontarget populations can cause interventions to have externalities and impact populations beyond the targeted beneficiaries. Recent experimental evidence documents substantial indirect benefits of interventions in various domains. For example, interventions to improve early childhood development generate spillover effects on neighboring children's skills and health through interactions among parents and children (Kusumawardhani, 2022; List et al., 2023), while medical treatments for children can create positive externalities for maternal mental health and the academic performance of their siblings (Daysal et al., 2022). Similarly, cash transfers to households can increase consumption among other village residents ineligible for such transfers (Angelucci & De Giorgi, 2009; Egger et al., 2022), and information disseminated through central community members improves village-wide vaccination rates (Banerjee et al., 2019). In agricultural communities, individuals in the social network of subsidized farmers experience spillovers in technology adoption, yields, and beliefs (Beaman et al., 2021; Carter et al., 2021; Vasilaky & Leonard, 2018). However, resource competition can also generate negative externalities, as research on agricultural training programs found that control households in areas with high treatment concentration experience adverse effects (Duflo et al., 2023).

Educational training programs offer an ideal setting to examine externality effects due to their inherently social nature. Research has shown that even brief social activities, such as interactions during freshman orientation, can create significant peer effects through network formation (Brade, 2024). This is because in schools and workplaces, regular interaction allows individuals to easily observe their peers' behavior. These social interactions fundamentally shape both skill acquisition and productivity development through multiple channels.

Research shows that individual productivity is significantly influenced by the characteristics and performance of coworkers—a phenomenon known as productivity spillovers (Battu et al., 2003; Cornelissen et al., 2017; Guryan et al., 2009). These spillovers can be both positive and negative: Knowledge sharing and skill complementarity create positive externalities (Azoulay et al., 2010; Jaravel et al., 2018), while competitive pressures can generate negative effects that limit individual growth and overall productivity (Breza et al., 2018; Cornelissen et al., 2017).

This study examines the spillover effects of a supplementary educational intervention for mathematics skills targeting boarding students in primary schools in rural China. Several characteristics of this setting make it ideal for studying how peer interactions facilitate spillover effects. First, boarding students reside on campus and rely primarily on school resources after class while interacting regularly with nonboarding peers during school hours. Additionally, primary schools in rural China feature fixed classroom assignments where students maintain consistent seating throughout the semester as teachers rotate between rooms. This structure facilitates stable peer networks and enables information diffusion between boarding and nonboarding students. Our cluster-randomized controlled trial spans 130 primary schools and includes over 3,000 boarding students across 352 classrooms in fourth through sixth grade. The experiment randomly assigned boarding students to one of three conditions: (i) computer-assisted learning (CAL), (ii) traditional workbook instruction (pencil-and-paper learning treatment), and (iii) a control group. While Ma et al. (2024) did not estimate the CAL or the workbook program to positively affect the test scores of boarding students, our study examines how these targeted interventions indirectly affect their 6,000 nonboarding classmates. This design addresses common challenges in estimating causal peer effects, such as simultaneity, correlated unobservables, and endogenous peer group formation (Manski, 1995). By introducing exogenous variation through targeted interventions while leveraging naturally formed social networks, we can identify genuine spillover effects.

The findings reveal distinct spillover effects from the two interventions. The CAL program had no significant impact on nonboarding students, but the workbook intervention generated notable negative spillovers for those who frequently interacted with treated boarding peers. These spillover effects intensified based on three key factors: the amount of time treated peers spent on the intervention, the total number of peers receiving treatment, and the social proximity between students as measured by their listing order in study partner surveys. Notably, the effects did not diminish with greater physical distance between students, suggesting that social proximity matters more than physical proximity in determining spillover intensity. Furthermore, the workbook intervention influenced broader classroom dynamics that extended beyond immediate peer connections, indicating that targeted educational programs can reshape entire classroom social networks rather than merely affecting direct peer relationships.

We then examine the mechanisms behind these effects, focusing on classroom climate dynamics and students' motivation for learning mathematics. According to Ryan and Deci (2017), motivation can be intrinsic (driven by interest) or instrumental (driven by future goals such as career success) (Eccles & Wigfield, 2002; Gagné & Deci, 2005; Simons et al., 2004). Our analysis reveals that the workbook intervention specifically undermined instrumental motivation: Nonboarding students exposed to treated peers became less confident that academic effort would yield long-term benefits. This effect was particularly pronounced among female students. In contrast, the CAL program, which was conducted outside the classroom setting, did not reduce motivation—highlighting the critical role of intervention visibility in shaping students' perceptions and responses to peer-targeted educational programs. Furthermore, we find no evidence of changes in competitive or cooperative behaviors between students, suggesting that motivational channels rather than behavioral competition drive the observed spillover effects.

Our study also reveals two main effects on teacher behavior. First, while classroom time use and assistance remained unchanged, the workbook intervention introduced evaluation bias: teachers overestimated the performance of students connected to treated peers and underestimated those without such ties. This aligns with the halo effect (Nisbett & Wilson, 1977; Thorndike, 1920) and findings from Ma et al. (2024) showing that teachers inflated their rankings of the treated students. Second, contrary to expectations, we found no evidence that the workbook intervention increased classroom competition. Nontarget students remained willing to collaborate, and test score dispersion within groups did not rise (Abramo et al., 2012; Fallucchi et al., 2021; Hoxby, 2000). Thus, the workbook program influenced motivation and perception but did not disrupt cooperative dynamics or resource allocation.

This study contributes to the literature on human capital development and intervention externalities in several ways. Prior research presents mixed evidence on spillover effects: Some document positive impacts on nontarget children (Islam et al., 2021; Kusumawardhani, 2022; List et al., 2023), while others find null or negative effects (Becker et al., 2022; Pedersen et al., 2017). We add to this debate by examining spillovers in a resource-constrained context in rural China where students often compete for limited opportunities, demonstrating the importance of minimizing perceived inequality when implementing interventions. Our findings related to the workbook intervention align with Angelucci et al. (2019), who showed that incentivizing peers can lead to negative spillovers due to perceived unfairness. We further compare two types of interventions (CAL versus workbook), and we adopt a longitudinal approach based on Abramitzky et al. (2021) to track how spillover effects evolve with social dynamics over time. This allows for a more nuanced understanding of both short- and long-term consequences of targeted educational interventions.

The study also contributes to the literature on peer effects in the workplace, with implications for organizational management and training program design. Building on Cornelissen et al. (2017) and Bentsen et al. (2019), who found modest peer effects on wages, we show that targeted interventions can reshape peer networks and change perceived status to influence confidence and development outcomes—similar to findings by Banerjee et al. (2024) in credit markets. Unlike Duflo et al. (2023), who focused on agricultural resource competition, we highlight how training can shift teacher expectations and peer dynamics. These results offer insights for managing equity and motivation when providing targeted support for individuals. On training design, our findings challenge the common approach of training a few individuals and relying on informal diffusion (Banerjee et al., 2019; Beaman et al., 2021; Chandrasekhar et al., 2022). While efficient, this strategy may demotivate untrained peers in competitive settings. This aligns with Duflo et al. (2023) but provides a more detailed view of the mechanisms. Our results also complement Anwar et al. (2024), who show that peer composition shapes training effectiveness. Overall, we show how targeted interventions can create unintended consequences, underscoring the need for careful program design in resource-limited environments.

2 Background

2.1 The School Environment in Rural China

Primary education in rural China offers a distinctive setting for studying student networks and resource access. Unlike many Western systems, students are assigned a fixed seating in classrooms while teachers rotated, which fosters stable peer groups while also limiting interactions beyond classmates. Another key feature is the boarding system, which creates two student populations that share the same classrooms but separate after school: Boarding students remain on campus and access school resources after class, while nonboarding students return home daily. This arrangement not only allows for implementing targeted interventions like CAL and workbook programs for boarding students during after-school hours, but also makes it an ideal context to examine peer effects, resource disparities, and group dynamics within shared classrooms.

2.2 Peer-to-Peer Information Diffusion in Academic Settings

Information diffusion among students involves both direct communication and observational learning. Students explicitly share knowledge through discussion and explanation (Webb, 1989), while also passively adopting strategies by observing successful peers, as described in the social learning theory (Bandura & Adams, 1977). These mechanisms can lead to the spread of effective study habits within peer groups (Alatas et al., 2016; Noriega-Campero et al., 2018). Collaborative settings like study groups further amplify information diffusion through shared problem-solving. However, the strength of social ties shapes the depth and frequency of exchange, such that stronger ties promote more substantive sharing (Gee, Jones, & Burke, 2017; Gee, Jones, et al., 2017; Granovetter, 1973). The social ties between boarding students who participate in the interventions and nonboarding students in the same classroom allow for examining whether and how the effect of the interventions diffuse via peer networks.

3 Experimental Design and Intervention

3.1 Experimental Design

We conducted a field experiment in rural primary schools in Shaanxi Province in northwestern China to evaluate the effectiveness of targeted supplementary education and potential spillover effects. A total of 130 schools across nine counties with low average income participated. From each school, one class with at least four boarding students was randomly selected from the fourth, fifth, and sixth grade, respectively. This yields a sample of 352 classes, each representing a unique school-grade combination. As shown in Figure B.1, the study had four phases: (1) a baseline survey of students, teachers, and principals in October 2017; (2) random assignment of the classes to treatment conditions within the 27 strata (nine counties $\times$ three grades); (3) random assignment of the 4,024 boarding students to CAL (1,345), workbook (1,289), or control (1,390) groups in November 2017; and (4) a follow-up survey in June 2018. Appendix Table A.1 summarizes the randomization.

Although nonboarding students did not receive training, we surveyed 6,414 of them who shared classrooms with boarding students: 2,061 in classes assigned to CAL, 2,093 in classes assigned to workbook, and 2,260 in classes assigned to control.

3.2 Intervention

The randomized controlled trial (RCT) assigned boarding students to one of three groups: (1) computer-assisted learning (CAL) with online math tutoring, (2) traditional paper-based workbooks, or (3) a control group. Only boarding students received the interventions, as they stayed on campus after class and could access school resources. CAL students used animated math games in the computer lab for 40 minutes each Sunday, while workbook students completed equivalent paper exercises in their classrooms. Nonboarding students were not directly treated, and teachers remained unaware of intervention details to minimize bias.

Ma et al. (2024) found that, on average, neither of the interventions affected the math test scores for boarding students, but both improved teacher-assigned grades (see Appendix Table A.2). The design allows us to compare spillover effects on untreated classmates. While EdTech may spark peer interest, barriers like device access and lack of in-person sharing may limit spillovers (Alajmi & Al-Qallaf, 2022). Traditional workbooks, by contrast, may diffuse more easily through peer interaction. Comparing the two interventions helps reveal how format and delivery influence peer spillover in educational settings.

4 Data

4.1 Math Test

Since the interventions focus on mathematics skills, we use math scores to measure the students' academic outcomes. Students took a 35-minute standardized exam in math at the baseline and the follow-up. The exam was grade-appropriate, tailored to the national and provincial-level mathematics curriculum, and constructed by trained psychometricians using a multi-stage process. We standardized math scores at the baseline and the follow-up into z-scores using each grade's mean and standard deviation. Figure B.2 shows the distribution of test scores for boarding and nonboarding students, respectively. While there are subtle differences in the score distributions, the overall patterns are quite similar at baseline. Nonboarding students demonstrate a slight advantage on average, as evidenced by their score distribution being marginally to the right of that of the boarding students.

4.2 Mediators

In resource-constrained educational environments, intense academic competition—driven by parental and teacher emphasis on rankings—can create a zero-sum mindset among students, reducing peer support and fostering sabotage (Drago & Garvey, 1998; Lazear, 1989). Empirical evidence from Chinese universities shows that such competition diminishes cooperation and increases conflict among students Chen and Hu (2024). Supplementary training programs may worsen these dynamics by increasing pressure, inequality, and demotivation among untreated peers (Ashcraft, 2002; Buunk & Ybema, 2003; Ramirez et al., 2016; Ryan & Deci, 2017), or alternatively, promote collaboration and motivation through role modeling or competitive inspiration (Jackson & Bruegmann, 2009). Additionally, teacher perceptions may shift due to the "Halo Effect" Thorndike (1920), where students associated with trained peers are evaluated more favorably—or unfavorably—based on proximity to the intervention (Alesina et al., 2024; Carlana, 2019; Lavy et al., 2012; Ma et al., 2024; Malouff et al., 2013).

Therefore, to explore the mechanisms underlying student performance, we examined student self-reported attitudes and teacher-observed behaviors at baseline and in the follow-up survey. To assess student attitudes, we used four validated scales from the 2012 Programme for International Student Assessment OECD (2014), which are widely applied in educational research (Lee, 2009; Pitsia et al., 2017; Thien et al., 2015). The scale includes 18 items describing math anxiety (5 items), academic confidence (5 items), intrinsic motivation (4 items), and instrumental motivation (4 items). Students indicated their agreement with each item on a four-point scale that ranged from "strongly disagree" to "strongly agree. Using the GLS weighting procedure described in Anderson (2008), We condensed these responses into four composite measures, with higher scores indicating greater levels of anxiety for each construct. We also assessed students' willingness to collaborate with a single item: "Do you like studying in a group?". Detailed scale items are available in Appendix C. To complement student self-reports, we collected teacher evaluations of students' engagement with math classes. Teachers reported the average numbers of times per week a student appeared distracted, how often they interrupted classmates, how often they utilize their full potential in assignments, and the duration of teacher assistance received. This combination of self-reported attitudes and teacher observations provides a comprehensive view of factors potentially influencing student academic performance.

4.3 Study Partner List

To evaluate interactions between nonboarding students and their boarding peers, we utilized study partner lists collected during both baseline and follow-up surveys (see Figure B.3). Each student was asked to identify up to ten classmates from the same classroom with whom they most frequently studied or discussed math. From these lists, we constructed a study group for each surveyed student comprising all the classmates they nominated. These data allow us to track changes in social networks between baseline and follow-up surveys and identify pairwise links between boarding and nonboarding students. Our data shows that nonboarding students reported an average of five study partners at baseline and six in the follow-up survey, with approximately 25% of them being boarding students.

4.4 Distance between Students

There is ample evidence that when students are physically close to one another in an academic context, they can influence each other's academic performance (Marmaros & Sacerdote, 2006; Rivera et al., 2010). Considering this, we tested whether the spillover effects of the interventions are stronger among the nonboarding students closer to their boarding peers, and whether the effects diminished with increasing distance. We leveraged graphical representations of the seating arrangements of each sample classroom, recorded by enumerators as the final segment of the baseline survey. Utilizing the data derived from this distribution in Figure B.4, we computed the relative distances between pairs of students through a series of calculations: First, we established a coordinate plane from the seating distribution, with the first row of desks as the horizontal axis and the first column as the vertical axis. The distance between adjacent desks on each axis was standardized as one desk. Subsequently, we assigned coordinates (a_i, b_j) to students based on their seat order relative to the origin, defined as the first seat in the first row. This allows us to determine the distance between any two seats with coordinates (a_{i1}, b_{j1}) and (a_{i2}, b_{j2}).

4.5 Socioeconomic Status of Student

We included the nonboarding students' and their teachers' socioeconomic and demographic characteristics, drawn from the baseline surveys, as covariates. Student variables included gender, parental education, family size, and relative household wealth. We measured household wealth by recording whether the family owned items including computers, internet devices, bicycles, microwaves, refrigerators, and air conditioners and standardizing the sum of these binary responses to construct a family asset index. Teacher variables included age, gender, and education level. Table A.3 presents student and household demographic characteristics and baseline math scores. Table 1 compares characteristics between the non-boarding peers in treatment and control classrooms using OLS regressions, with treatment status as the dependent variable and student, household, and classroom factors as covariates. We control for class size, number of boarding peers, and include strata fixed effects, with robust standard errors clustered at the class level. Results reveal imbalances at baseline: On average, classes assigned to the workbook condition had more boarding students, and classes assigned to the CAL condition were more likely to have a male math teacher. These covariates are included in our main models to isolate exogenous spillover effects.

5 Empirical Approach

One primary goal of this study is to estimate the spillover effect of two interventions on nontarget student academic performance. To do so, we apply a linear regression model to the math scores of nonboarding students:

$$\begin{aligned}\text{Score}_{ic,\text{endline}} = \beta_0 &+ \beta_1 \text{CAL}_{c,\text{baseline}} + \beta_2 \text{Workbook}_{c,\text{baseline}} \\ &+ \beta_3 \text{Boardingnum}_{c,\text{baseline}} + \beta_4 \text{Classsize}_{c,\text{baseline}} + \beta_5 X_{ic,\text{baseline}} \\ &+ \beta_6 \text{Score}_{ic,\text{baseline}} + \pi_s + \epsilon_{ic}\end{aligned} \tag{1}$$

where $Score_{ic,endline}$ is the standardized mathematics score of $student_{ic}$ at the time of the follow-up survey; $CAL_{c,baseline}$ and $Workbook_{c,baseline}$ represent the treatment status of $class_c$; $Boardingnum_{c,baseline}$ is the number of boarding students in $class_c$ at baseline; $Classsize_{c,baseline}$ is the size of $class_c$; and $X_{ic,baseline}$ is a vector of controls at baseline survey of $student_{ic}$, including student characteristics listed in in Table 1. $Score_{ic,baseline}$ is the standardized mathematics score of studentic at the time of the baseline survey. π_s is a set of county-grade (strata) fixed effects, with 27 county-grades (9 counties and 3 grades). ϵ_{ic} is a random error term. Because including classroom fixed effects could introduce exclusion bias in peer effect estimation(Caeyers & Fafchamps, 2024), we instead adjusted for standard errors at the classroom level while controlling for observable classroom characteristics. This approach accounts for within-classroom error correlation without introducing the mechanical negative relationship between an individual's outcome and their peers' outcomes that can arise with fixed effects. We also estimated Equation (1) separately for nonboarding students with boarding peers and those without boarding peers. This approach isolates the potential influence of direct connections to treated students.

6 Main Results

6.1 Spillover Effect on Academic Performance

Table 2 presents the results of estimating Equation (1). The CAL intervention had no spillover impact on nonboarding students. On the other hand, the workbook intervention had a discernible adverse spillover effect: Nonboarding students in classes assigned to the workbook intervention experienced a decrease in math test scores by 0.087 standard deviations compared to nonboarding students in the control classes after one academic year, significant at the 5% level. To assess the robustness of our findings, we conducted additional analyses using alternative measures of academic performance. We replaced absolute scores with two types of relative rankings: within-class rankings, which are typically of greatest interest to teachers and parents, and within-study group rankings, which reflect relative academic performance in a student's immediate social network. The results of these analyses, presented in Table A.4 (for within-class rankings) and Table A.5 (for within-study group rankings), demonstrate the consistency of our results in Table 2.[1]

The negative spillover effects experienced by nonboarding students are most likely explained by their interactions with boarding students. This is because nonboarding students did not directly receive the interventions; The spillover effects cannot be explained by a

[1]As an additional robustness check, we removed 535 students (5.2% of the sample) who switched from nonboarding to boarding status during the second semester. Results remained consistent with our main findings, as shown in Table A.6.

relative performance decline, either, since neither intervention significantly improved the boarding students' academic performance. As such, the most likely explanation is that social interactions between boarding and nonboarding students led to the spillover effects, with the workbook intervention having a stronger negative effect on the nonboarding students' math scores (-0.087 SD) than the CAL program.

To test this conjecture, we perform an additional analysis of peer effects between nonboarding students and their self-identified boarding peers within the same classroom (see Table A.7). This analysis reveals that under the control condition, nonboarding students benefit from high-performing boarding peers, with a one standard deviation increase in a boarding student's score associated with a 0.083 standard deviation increase in their paired nonboarding peer's performance. However, the treatment interventions fundamentally alter this relationship. While the CAL program maintains the cooperative dynamic, the workbook treatment creates a more competitive environment. The significant interaction term (0.067, $p < 0.05$) combined with the negative main treatment effect (-0.096, $p < 0.05$) indicates that workbook-treated boarding students' gains are partially offset by negative effects on their nonboarding peers, consistent with zero-sum competitive dynamics.

6.2 Intensity of Treatment

To test the robustness of our findings, we replaced binary treatment indicators with continuous measures of treatment intensity, defined as the average time boarding students spent on CAL or workbook activities (divided into deciles). The actual average exposure time—590 minutes for CAL and 636 minutes for the workbook program—fell short of the planned 1,120 minutes due to technical issues, student absences, and school disruptions. Because prior research suggests that higher treatment intensity may lead to stronger peer spillovers (Babcock & Hartman, 2010), we explored whether spillover effects increased with exposure. Regression results in Table 3 show that an increase in the time boarding students spend on workbooks is associated with a significant decrease in the math scores of nonboarding students, particularly those with boarding peers. Specifically, a 10% increase in workbook time is associated with a 0.014 standard deviation decline in the academic performance of non-boarding peers, and a 0.018 standard deviation decline among non-boarding students who have treated peers. Figure B.5 further illustrates these effects across quintiles. The workbook intervention exhibits similar pattern: greater time spent by treated peers corresponds to stronger negative spillover effects on non-boarding students. The CAL intervention shows no such relationship. These results support a dose-response relationship in peer spillovers, though given the non-random nature of treatment intensity, we interpret these findings as suggestive rather than causal.

One possible reason for the difference in spillovers between the two interventions is the environment in which they were conducted. The CAL program took place in a computer lab, where information diffusion is less effective compared to the classroom setting used for the workbook treatment. In classrooms, students are in closer proximity, which facilitates information sharing. Additionally, the implementation process may have played a significant role. Students in the CAL program only had access to the material during the designated weekly session in the computer labs. On the other hand, although students in the workbook program were asked to turn in their workbooks after each weekly session, some students retained their workbooks to finish the exercises outside of the session. This can make the additional training they receive from the intervention more salient to their nonboarding peers. The paired-student analysis supports this interpretation, as the workbook treatment specifically disrupts the typically positive peer learning dynamics and transformed collaborative relationships into competitive ones.

6.3 Number and Order of Boarding Peers

To explore how network characteristics influence spillover effects, we modified our analysis to account for both network density and connection strength. Network density is proxied by number of treated peers, and we expect that spillover effects are amplified when more peers in a student's network receive intervention (Centola, 2010). Simultaneously, stronger connections between individuals, typically characterized by more frequent interactions, may magnify spillover effects by providing increased opportunities for information and behavior transmission (Granovetter, 1973). To capture these dynamics, we introduced interactions between treatment assignment and the network variables into Equation (1). These network variables include the number of boarding peers and whether the first nominated peer is a boarding student. Although students were not explicitly instructed to nominate study partners based on interaction frequency, we assume that peers nominated earlier in the list interact more frequently with the respondent. This analysis focuses on nonboarding students who nominated at least one study partner and is conducted only for those with boarding peers. By examining these interactions, we aim to understand how network structure and connection strength modulate the spillover effects of our educational interventions. The Equation is as following:

$$\begin{aligned}
\text{Score}_{ic,\text{endline}} = \beta_0 &+ \beta_1 \text{CAL}_{c,\text{baseline}} + \beta_2 \text{Workbook}_{c,\text{baseline}} \\
&+ \beta_3 \text{Network}_{ijc,\text{baseline}} + \beta_4 (\text{CAL}_{c,\text{baseline}} \times \text{Network}_{ijc,\text{baseline}}) \\
&+ \beta_5 (\text{Workbook}_{c,\text{baseline}} \times \text{Network}_{ijc,\text{baseline}}) \\
&+ \beta_6 \text{Boardingnum}_{ic,\text{baseline}} + \beta_7 \text{Classsize}_{c,\text{baseline}} \\
&+ \beta_8 X_{ic,\text{baseline}} + \beta_9 \text{Score}_{ic,\text{baseline}} + \pi_s + \epsilon_{ic}
\end{aligned} \tag{2}$$

where $Network_{ijc,baseline}$ denotes network-related variables for $student_{ic}$ and all other items are consistent with Equation (1). To ensure that treatment coefficients represent average spillover effects, we de-meaned the nonboarding students' number of boarding peers and distance with their boarding peers before creating interaction terms. Our analysis reveals that network characteristics significantly moderate the spillover effects of educational interventions, as shown in Table 4. For the workbook intervention, we find that for each additional treated peer, the negative spillover effect increases by 0.031 standard deviations

(SD) ($p < 0.10$), and when the first-nominated peer is a boarding student (treated), the negative spillover effect is 0.120 SD larger ($p < 0.05$). In contrast, the spillover effects of the CAL program did not vary significantly with the number of treated peers or whether the first-nominated peer is a treated student. These findings can be interpreted through social comparison theory: Stronger ties facilitate greater information exchange (Levin & Cross, 2004) making frequently-interacting peers more relevant comparators (Zell & Alicke, 2010). As such, students with more treated peers may perceive not receiving treatment as a disadvantage to their own educational opportunities (Fiske, 2010). The differential effects between workbook and CAL interventions suggest that the visibility of the workbooks outside the treatment period may play a role in triggering social comparison processes. These results underscore the importance of considering network structures and social dynamics when implementing and evaluating educational interventions, as they can significantly influence spillover effects.

6.4 Spatial Patterns in Spillover Effects of Interventions

The relative distance between treated peers, represented by path length in network terms, may influence the persistence of peer effects as they spread through the network. Since the spread of information and the strength of peer influence declines with spatial distance (List et al., 2023), we expect that the spillover effects of the workbook intervention should decrease as the physical distance between nonboarding students and their boarding peers increases. To test this, we calculated the mean value of desk distance to boarding peers, ranging from 0 to 12, with values of 9-12 recoded as 8 to ensure adequate statistical power. This analysis only includes samples of individuals who have interacted with boarding peers at baseline.

We examine heterogeneity by distance using Equation (2), where cross terms represent interactions between treatment assignment and relative distance. As shown in Table 4, we did not observe a significant difference in spillover effects as desk distance to a treated peer increased. In other words, we found no evidence that the spillover effects of either intervention were only localized and decreased as the distance to a treated peer increased. Several factors could explain this result: For example, the fixed classroom assignment facilitates interaction among classmates regardless of seating distance, the high visibility of workbooks makes them salient to all students, and effects may also spread through social networks rather than physical proximity. Additionally, the observed spillover effect on student motivation may operate at a classroom-wide level through general awareness rather than direct peer transmission. This suggests that in close-knit classroom environments, the visibility and psychological impact of targeted interventions may cause them to have broader spillover effects than on those physically proximal to the treated students.

7 Identifying the Social Network Mechanism

Extensive research in education economics has established that peer effects and teacher-student interactions constitute fundamental channels through which educational interventions transmit across classrooms. The peer effects literature demonstrates that students' academic outcomes are significantly influenced by their classmates' performance through both competitive and collaborative mechanisms (Jackson & Bruegmann, 2009). Simultaneously, teacher behavior research reveals that educators' perceptions and expectations substantially affect student achievement through differential attention allocation and evaluation biases (Thorndike, 1920). However, when targeted interventions create performance disparities within classrooms, these mechanisms may generate unintended spillover effects that alter classroom dynamics in ways that either reinforce or undermine the original intervention's effectiveness.

In educational markets characterized by relative performance evaluation, students face strategic interactions where individual academic outcomes depend on both absolute achievement and relative standing within peer groups. Standard tournament theory suggests that supplementary training for select students intensifies competition by increasing performance variance, generating negative externalities through direct competition effects where students perceive academic success as a zero-sum game (Drago & Garvey, 1998; Lazear, 1989). Psychological stress mechanisms may also activate when students engage in upward social comparisons, leading to increased anxiety and reduced cognitive performance (Ashcraft, 2002; Buunk & Ybema, 2003; Ramirez et al., 2016), while perceived inequity in resource allocation undermines intrinsic motivation as predicted by Self-Determination Theory (Ryan & Deci, 2017). Alternatively, peer effects literature suggests that successful students can generate positive externalities through knowledge transmission and motivational channels, potentially shifting classroom dynamics from competitive to collaborative equilibria (Jackson & Bruegmann, 2009). Recent evidence from Chinese universities confirms that competitive environments reduce cooperation and increase antisocial behaviors among peers Chen and Hu (2024), though the net effect of targeted interventions on classroom climate remains theoretically ambiguous.

To examine these competing theoretical predictions, we investigate three primary mechanisms through which interventions may generate spillover effects: classroom competition dynamics, student psychological responses, and teacher behavioral changes. We re-estimated Equation (1) using various mediators from the follow-up survey—instead of test scores—as outcomes, including peer relationships, student-teacher interactions, and academic self-perception measures. We stratify results by whether nonboarding students had baseline ties to boarding peers, allowing us to isolate effects transmitted through direct social connections versus general classroom dynamics.

7.1 Competition Change in the Classroom

We first tested the competition hypothesis by examining whether supplementary education interventions created a more competitive learning environment. Following Abramo et al. (2012), Fallucchi et al. (2021), and Hoxby (2000), we analyzed the impact of interventions on non-boarding students' willingness to work in groups, test score dispersion within classes or study groups[2], and the likelihood of having boarding peers at follow-up. Our results in Table 5 revealed no negative spillover effects from the workbook intervention on these measures for non-boarding students with boarding peers. Contrarily, the Computer-Assisted Learning (CAL) intervention increased these students' likelihood of studying in groups. For non-boarding students without links to treated students, the workbook intervention also increased their willingness to study in groups but had no significant effect on test score dispersion or number of boarding peers. These findings consistently indicate that the supplementary education interventions did not negatively impact the collaborative atmosphere or significantly increase competitiveness within the classroom, suggesting that increased competition is unlikely to be the mechanism behind the observed negative spillover effects on non-boarding students.

7.2 Academic Anxiety and Motivation of Nonboarding Students

To investigate psychological channels, we employed Equation (1) replacing outcomes with mediators related to math learning processes, including academic anxiety, confidence, and both intrinsic and instrumental motivation. The academic anxiety measure is negatively coded, meaning that higher scores indicate greater anxiety, while confidence and motivation measures are positively coded. Table 6 presents results revealing that workbook intervention effects on nonboarding students' anxiety and motivation vary depending on their interaction level with treated boarding students. For nonboarding students who study alongside boarding peers, the workbook intervention reduces their instrumental motivation by 0.051 points ($p < 0.10$). Conversely, no negative spillover effects on academic motivation were observed for nonboarding students lacking direct interaction with boarding peers. Interestingly, the workbook intervention reduces anxiety scores by 0.1 points ($p < 0.05$) among non-boarding students with limited interaction with boarding peers, suggesting unexpected stress alleviation.

[2]Test score dispersion is measured by the standard deviation of test scores within each study group or classroom.

Table 7 reveals that nonboarding students with boarding peers experienced significant negative spillover effects in two instrumental motivation items: "Making some efforts in mathematics is worthwhile, as it will be helpful to me in the future" (-0.042 points, $p < 0.05$) and "Studying mathematics is very important to me because it will enhance my future job skills" (-0.041 points, $p < 0.10$). These effects indicate that interventions may alter nonparticipating students' perceptions of mathematical learning's long-term value, potentially due to perceived educational deprioritization, questioning of regular classroom instruction value, or reassessment of academic standing relative to peers receiving additional support.

7.3 Teacher Perceptions

To examine teacher behavioral mechanisms, we investigated whether the Halo Effect Theory (Thorndike, 1920) could generate spillover effects through altered teacher perceptions. Building on Ma et al. (2024)'s observation that interventions improved teachers' rankings of boarding students, we hypothesized that supplementary training serves as a positive signal to teachers, potentially creating evaluation biases affecting teacher-student interactions. We measured teacher evaluation bias by comparing teachers' student ranking evaluations to actual score rankings within classrooms, employing this post-intervention bias as the outcome variable in Equation (1) while controlling for baseline bias.

Table 8 presents analysis revealing significant workbook intervention effects on teacher perceptions associated with students' boarding peer connections. For nonboarding students with boarding peers, teachers tend to overestimate academic performance, ranking these students approximately 1 position higher ($p < 0.05$) than actual test score rankings. Conversely, for nonboarding students without boarding peers, teachers tend to underestimate performance, ranking these students about 1.6 positions lower ($p < 0.05$) than actual rankings. These findings suggest the workbook intervention induces Halo Effects, creating positive bias for nonboarding students interacting with intervention recipients while generating negative bias for those without such connections.

Table 9 examines whether this bias affected teacher time allocation and student management through post-intervention assessments of nonboarding students' in-class behavior and assistance levels. Analysis reveals no negative spillover effects from interventions. Notably, for nonboarding students without boarding peers, workbook treatment leads teachers to perceive increased likelihood (6.8%, $p < 0.1$) that students exert full effort studying mathematics. While interventions led teachers to overestimate academic performance of students connected with treated peers, they did not significantly alter evaluations of other behaviors or reduce time investment, indicating that interventions primarily influenced academic assessment perceptions without affecting overall teacher approach to student interaction or classroom management.

To test whether spillover effects operate through direct peer-to-peer interactions rather than general classroom dynamics, we examine how individual peers' academic performance directly affects nonboarding students when only boarding peers receive treatment (see Table A.7). Analysis reveals that nonboarding students normally benefit from high-performing peers regardless of boarding status, with positive peer effects of similar magnitude. However, workbook treatment fundamentally disrupts peer relationships throughout classroom social networks, creating negative effects not only when boarding peers receive treatment but also affecting relationships among nonboarding peers who received no treatment themselves. This suggests workbook interventions generate classroom-wide competitive dynamics extending beyond direct treatment recipients to alter the entire peer learning ecosystem, shifting collaborative norms toward competition. In contrast, CAL treatment maintains cooperative peer learning dynamics across all student pairs.

8 Heterogeneity on Gender

In rural Chinese primary schools, educational disparities between female and male students naturally arise from gender discrimination in the social environment and gender inequality in family values (Hannum, 2005; Hannum et al., 2021; Wu, 2024). Therefore, it is worthwhile to investigate gender-specific responses to peers receiving supplementary education in the context of such an uneven distribution of educational resources. We analyze nonboarding students separately by gender using Equation (1). Panel A of Table 10 reveals that workbook interventions have a more pronounced adverse spillover effect on female students who have boarding peers, decreasing their scores by 0.108 SD ($p < 0.01$), while showing no significant impact on male students. We also estimated spillover effects by different genders for nonboarding students without boarding peers, presented in Panel B. These findings indicate that spillover effects from both interventions are insignificant when nonboarding students are not exposed to boarding peers, suggesting that negative spillovers can be mitigated in such circumstances.

We further analyzed whether instrumental motivation remains the primary channel through which negative spillover operates, using Equation (1) but replacing the outcome variable with instrumental motivation and its first two items, which have been significantly impacted by the workbook intervention for the entire population. Analyzing separately the subgroups of nonboarding students with and without boarding peers, we found that for nonboarding students with boarding peers, the workbook intervention reduced female students' instrumental motivation by 0.119 points ($p < 0.05$). Specifically, it affected two items: "Studying mathematics is very important to me because it will enhance my future job skills" and "Mathematics is a relatively important subject for me because I will need to use it in my future studies." In contrast, the workbook intervention had no significant effect on instrumental motivation for nonboarding students without boarding peers. According to Table A.8, when comparing the above treatment effects by gender, we find significant differences between female and male students in the workbook treatment, but only among students with boarding peers (Panel A). Specifically, the difference is significant for instrumental motivation (0.121, $p < 0.05$) and math effort worthwhile (0.104, $p < 0.05$). In contrast, among students without boarding peers (Panel B), we observe no significant gender differences in treatment effects.

These results reinforce the presence of spillover effects between boarding and nonboarding peers, with future-oriented motivation emerging as a key mechanism. Notably, even when workbook interventions are equally distributed among boarding students, nonboarding female students may feel at a greater disadvantage than males, who are more accustomed to receiving educational advantages. This perceived disparity can heighten female students' sensitivity to resource gaps, potentially undermining their motivation if they view the interventions as essential for success.

9 Conclusion

Our findings reveal distinct patterns in how supplementary educational interventions affect nontarget students. Traditional workbook interventions negatively impact peers who don't receive the intervention, while education technology interventions conducted outside the classroom show no spillover effects. These negative spillovers appear linked to direct classroom interactions with intervention recipients. The magnitude of negative effects increases with both "treatment dose" (time spent on workbooks) and exposure to treated peers. Specifically, the spillover intensity correlates with both the number of treated peers and the frequency of interactions with them. These effects manifest primarily through decreased future-oriented motivation among nontarget students. Importantly, neither intervention altered teachers' attention to nonboarding students or classroom competition levels. In stead, we observed shifts in teacher perceptions, with teachers showing preference for nontarget students who frequently study with their treated peers. In resource-constrained environments, boys - who typically receive more familial attention - demonstrate greater resilience to these negative spillovers.

Our findings carry several important policy implications. First, in designing educational interventions, policymakers must consider not only direct effects on targeted students but also potential spillovers on their peers. The contrast between the null effects of EdTech and the negative spillovers from workbook interventions highlights how delivery methods and context matter. Second, in organizational settings, targeted training or resource allocation should be assessed for its impact on group dynamics and motivation. The observed shifts in teacher perceptions and reduced future-oriented motivation among non-targeted students suggest the need to preserve perceived equity. Third, for human capital development, our results challenge the assumption that selective interventions in low-resource settings benefit all. Instead, they underscore the importance of inclusive designs or complementary measures to minimize negative externalities. These considerations are especially critical where resources are limited and equity directly shapes long-term development and mobility.

Our study has several limitations. First, 535 students (5.2%) changed from nonboarding to boarding during the second semester. Although they did not receive the intervention, their learning environment shifted. We excluded these students and confirmed through robustness checks that their exclusion did not alter our main results (Table A.5). Second, information can spread beyond study groups or classrooms, including peers from other classes or neighborhoods. However, lacking a full map of students' social networks, we cannot capture spillovers through these broader channels—an area for future research.

Tables

Table 1 – Comparisons of Sample Characteristics Between Treatment and Control Groups

	(1) Workbook-Control	(2) CAL Program-Control
Student's gender (1=male; 0=female)	0.013	0.004
	(0.013)	(0.013)
Mother graduated from junior high school	0.012	0.021
	(0.017)	(0.019)
Father graduated from junior high school	-0.023	0.010
	(0.017)	(0.017)
Family asset index	-0.010	-0.005
	(0.006)	(0.005)
Number of family members	0.006	0.009
	(0.007)	(0.007)
Teacher's gender (1=male; 0=female)	0.026	0.190**
	(0.090)	(0.095)
Teacher's age (years)	0.004	-0.006
	(0.005)	(0.005)
Teacher graduated from two-year college	0.002	0.060
	(0.102)	(0.098)
Class size	-0.005	0.000
	(0.003)	(0.003)
Number of boarding peers	0.000	-0.007
	(0.007)	(0.007)
Standardized score at baseline survey	0.010	0.014
	(0.012)	(0.012)
Exposure to boarding peers	-0.055**	-0.005
	(0.025)	(0.024)
Constant	0.492**	0.581**
	(0.218)	(0.226)
Observations	4,353	4,320
R-squared	0.085	0.082
P-value (F-test)	0.382	0.353

Note: Cluster and class fixed effects are included; Standard errors are clustered at the grade level and robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2 – Treatment Effect on nonboarding Student Academic Performance

	(1) Whole Sample	(2) Has Boarding Peers	(3) No Boarding Peers
CAL Program	0.018	0.006	0.047
	(0.038)	(0.041)	(0.053)
Workbook Treatment	-0.048	-0.087**	0.014
	(0.038)	(0.041)	(0.053)
Control Mean	0.021	0.071	-0.063
Observations	6,414	3,898	2,516
R-squared	0.422	0.422	0.435

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level).
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3 – Effect of Treatment Intensity on nonboarding Student Academic Performance

	(1) Whole Sample	(2) Has Boarding Peers	(3) No Boarding Peers
CAL Program	-0.002	-0.002	0.000
	(0.006)	(0.006)	(0.008)
Workbook Treatment	-0.014**	-0.018***	-0.008
	(0.005)	(0.006)	(0.007)
Control Mean	0.021	0.071	-0.063
Observations	6,414	3,898	2,516
R-squared	0.423	0.422	0.435

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level).
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4 – Heterogeneity on Social Network Characteristics

	Dependent Variable: Math Test Score		
	(1) Number of Boarding Peers	(2) Order of Nominated Boarding Peers	(3) Desk Distance with Boarding Peers
CAL Program	0.017	0.038	0.005
	(0.038)	(0.044)	(0.041)
Workbook Treatment	-0.047	-0.018	-0.082**
	(0.037)	(0.040)	(0.041)
Number of Boarding Peers	0.019		
	(0.013)		
CAL*Number of Boarding Peers	-0.012		
	(0.018)		
Workbook*Number of Boarding Peers	-0.031*		
	(0.018)		
First Nominated Peer is Boarding		0.047	
		(0.041)	
CAL*First Nominated Peer is Boarding		-0.088	
		(0.057)	
Workbook*First Nominated Peer is Boarding		-0.120**	
		(0.053)	
Desk Distance with Boarding Peers			0.011
			(0.011)
CAL*Desk Distance with Boarding Peers			-0.016
			(0.015)
Workbook*Desk Distance with Boarding Peers			0.005
			(0.017)
Control Mean	0.021	0.029	0.071
Observations	6,414	6,343	3,894
R-squared	0.426	0.420	0.422

Note: Social network characteristics were collected from baseline, including number, order of boarding peers and distance with boarding peers. The number of boarding peers and distance with boarding peers have been de-meaned by nonboarding students. The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level). Missing values in column 2 occur because 71 students in our sample have no friends. We only include samples of whoever interacted with boarding peers in the regression in column 3.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5 – Treatment Effect on Non-boarding Student Cooperation and Competition

	(1) Like studying in a group?	(2) Test score dispersed in the group	(3) Test score dispersed in the class	(4) Number of boarding peers
Panel A: Has Boarding Peers				
CAL Program	0.053	-0.025	-0.023	0.014
	(0.028)	(0.020)	(0.018)	(0.174)
Workbook Treatment	0.036	-0.006	-0.001	-0.107
	(0.025)	(0.019)	(0.018)	(0.179)
Control Mean	0.596	0.780	0.898	6.008
Observations	3,892	3,897	3,898	3,898
R-squared	0.044	0.168	0.326	0.147
Panel B: No Boarding Peers				
CAL Program	0.043	-0.031	-0.002	-0.024
	(0.028)	(0.028)	(0.023)	(0.243)
Workbook Treatment	0.058	-0.013	0.009	-0.329
	(0.027)	(0.028)	(0.021)	(0.224)
Control Mean	0.548	0.758	0.893	5.640
Observations	2,512	2,424	2,516	2,516
R-squared	0.036	0.181	0.409	0.160

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level). Missing values in columns 1 and 2 are due to some students not answering the questions.

Table 6 – Treatment Effect on nonboarding Student Anxiety and Motivation

	(1) Anxiety	(2) Confidence	(3) Intrinsic Motivation	(4) Instrument Motivation
Panel A: Has Boarding Peers				
CAL Program	-0.002	0.019	0.017	-0.007
	(0.034)	(0.029)	(0.036)	(0.030)
Workbook Treatment	-0.020	0.011	0.004	-0.051*
	(0.033)	(0.029)	(0.034)	(0.031)
Control Mean	0.024	-0.019	-0.011	0.092
Observations	3,850	3,832	3,841	3,870
R-squared	0.231	0.338	0.226	0.138
Panel B: No Boarding Peers				
CAL Program	0.028	0.053	-0.007	0.013
	(0.048)	(0.046)	(0.047)	(0.048)
Workbook Treatment	-0.102**	0.049	-0.002	-0.027
	(0.050)	(0.046)	(0.042)	(0.043)
Control Mean	0.034	-0.018	-0.006	0.057
Observations	2,484	2,475	2,485	2,497
R-squared	0.232	0.298	0.241	0.132

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level). Missing values are due to some students not answering the questions.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 7 – Treatment Effect on nonboarding Student Motivation Items (Students with Boarding Peers Only)

	(1) Math effort worthwhile	(2) Math for future job	(3) Math for future studies	(4) Math knowledge for future job
CAL Program	-0.004	-0.001	0.001	0.001
	(0.022)	(0.022)	(0.022)	(0.023)
Workbook Treatment	-0.042**	-0.041*	-0.017	-0.024
	(0.021)	(0.023)	(0.020)	(0.023)
Control Mean	0.495	0.417	0.427	0.469
Observations	3,873	3,878	3,884	3,890
R-squared	0.093	0.070	0.068	0.074

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level). The observations include only nonboarding students who have boarding peers. Missing values are due to some students not answering the questions. Column headers represent abbreviated versions of the following statements: (1) Making some efforts in mathematics is worthwhile, as it will be helpful to me in the future; (2) Studying mathematics is very important to me because it will enhance my future job skills; (3) Mathematics is a relatively important subject for me because I will need to use it in my future studies; (4) I will learn a lot of knowledge from mathematics, and this knowledge will be helpful for me to find a job in the future.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8 – Treatment Effect on Evaluation of Ranking from Teachers Comparing to Actual Ranking

	(1) Whole Sample	(2) Has Boarding Peers	(3) No Boarding Peers
CAL Program	-0.247	0.036	-1.028
	(0.306)	(0.435)	(0.760)
Workbook Treatment	-0.019	0.993**	-1.602**
	(0.298)	(0.445)	(0.669)
Control Mean	0.508	-0.975	0.574
Observations	5,967	3,618	2,349
R-squared	0.597	0.615	0.579

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level). Missing values are due to a lack of teacher evaluations from some classes.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9 – Treatment Effect on Teacher's Evaluation of nonboarding Students' Class Performance and their Assistance

	(1) Distraction Frequency	(2) Interrupting Classmates	(3) Math Ability Utilization	(4) Help from Teachers (Min)
Panel A: Has Boarding Peers				
CAL Program	-0.004	0.007	-0.019	-0.699
	(0.038)	(0.028)	(0.034)	(1.702)
Workbook Treatment	0.048	0.049	-0.030	0.117
	(0.034)	(0.031)	(0.032)	(1.618)
Control Mean	0.534	0.295	0.589	17.644
Observations	3,814	3,813	3,812	3,763
R-squared	0.169	0.143	0.152	0.191
Panel B: No Boarding Peers				
CAL Program	0.054	-0.005	0.040	-0.473
	(0.041)	(0.037)	(0.042)	(1.722)
Workbook Treatment	-0.005	0.017	0.068*	0.749
	(0.039)	(0.042)	(0.039)	(1.775)
Control Mean	0.577	0.353	0.478	17.692
Observations	2,459	2,459	2,460	2,423
R-squared	0.185	0.142	0.152	0.215

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level). Missing values are due to a lack of teacher evaluations from some classes. Column headers represent: (1) Distraction Frequency at Class, (2) Frequency of Interrupting Classmates, (3) Frequency to Give Full Play to Mathematics Ability, (4) Getting Help from Teachers (Minutes).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 10 – Heterogeneity in Gender

	(1)		(2)		(3)		(4)	
	Math test score		Instrumental motivation		Math effort worthwhile		Math for future job	
	Female	Male	Female	Male	Female	Male	Female	Male
Panel A: Has boarding peers								
CAL Program	-0.019	0.021	-0.042	0.035	-0.018	0.003	-0.022	0.058
	(0.050)	(0.050)	(0.051)	(0.042)	(0.038)	(0.035)	(0.041)	(0.040)
Workbook	-0.108**	-0.073	-0.119**	-0.008	-0.098***	0.000	-0.070*	-0.045
	(0.050)	(0.052)	(0.049)	(0.043)	(0.036)	(0.033)	(0.041)	(0.043)
Control Mean	0.098	0.046	0.059	0.018	3.475	3.409	3.289	3.288
Observations	1,872	2,026	1,857	2,013	1,864	2,020	1,869	2,022
R-squared	0.408	0.442	0.147	0.137	0.117	0.091	0.071	0.079
Panel B: No boarding peers								
CAL Program	0.037	0.056	-0.013	0.032	-0.021	0.002	0.021	0.041
	(0.070)	(0.059)	(0.072)	(0.060)	(0.050)	(0.050)	(0.056)	(0.050)
Workbook	-0.074	0.081	-0.008	-0.063	-0.021	-0.058	0.032	-0.005
	(0.061)	(0.064)	(0.059)	(0.061)	(0.047)	(0.047)	(0.051)	(0.051)
Control Mean	-0.049	-0.074	0.042	-0.060	3.451	3.378	3.299	3.210
Observations	1,142	1,374	1,136	1,362	1,141	1,371	1,140	1,367
R-squared	0.434	0.457	0.135	0.144	0.110	0.097	0.079	0.073

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at class level). Missing values are due to some students not answering the questions. Dependent variables: (1) Standardized math test score, (2) Instrumental motivation, (3) "Making some efforts in mathematics is worthwhile, as it will be helpful to me in the future", (4) "I will learn a lot of knowledge from mathematics, and this knowledge will be helpful for me to find a job in the future".
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

References

Abramitzky, R., Lavy, V., & Pérez, S. (2021). The long-term spillover effects of changes in the return to schooling. *Journal of Public Economics*, *196*, 104369. https://doi.org/10.1016/j.jpubeco.2021.104369

Abramo, G., Cicero, T., & D'Angelo, C. A. (2012). The dispersion of research performance within and between universities as a potential indicator of the competitive intensity in higher education systems. *Journal of Informetrics*, *6*(2), 155–168. https://doi.org/10.1016/j.joi.2011.11.007

Alajmi, B. M., & Al-Qallaf, C. L. (2022). Fostering knowledge-sharing behavior through social capital: The implications of face-to-face and online interactions. *Global Knowledge, Memory and Communication*, *71*(4), 274–292. https://doi.org/10.1108/GKMC-01-2021-0007

Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R., & Olken, B. A. (2016). Network structure and the aggregation of information: Theory and evidence from indonesia [Publisher: American Economic Association]. *The American Economic Review*, *106*(7), 1663–1704. Retrieved September 30, 2024, from https://www.jstor.org/stable/43861109

Alesina, A., Carlana, M., La Ferrara, E., & Pinotti, P. (2024). Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review*, *114*(7), 1916–1948. https://doi.org/10.1257/aer.20191184

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495. https://doi.org/10.1198/016214508000000841

Angelucci, M., & De Giorgi, G. (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? [Publisher: American Economic Association]. *The American Economic Review*, *99*(1), 486–508. Retrieved November 18, 2024, from https://www.jstor.org/stable/29730193

Angelucci, M., Prina, S., Royer, H., & Samek, A. (2019). Incentives and unintended consequences: Spillover effects in food choice [Publisher: American Economic Association]. *American Economic Journal: Economic Policy*, *11*(4), 66–95. https://doi.org/10.1257/pol.20170588

Anwar, S., Baird, M., Engberg, J., & Smart, R. (2024). Do sectoral training programs reduce arrests?: Evidence from a low-income targeted training program RCT. *Journal of Human Resources*, 1122–12629R4. https://doi.org/10.3368/jhr.1122-12629R4

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *11*(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Azoulay, P., Zivin, J. S. G., & Wang, J. (2010). SUPERSTAR EXTINCTION. *QUARTERLY JOURNAL OF ECONOMICS*, *125*(2), 549–589.

Babcock, P., & Hartman, J. (2010, December). *Networks and workouts: Treatment size and status specific peer effects in a randomized field experiment* (w16581). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w16581

Bandura, A., & Adams, N. E. (1977). Analysis of self-efficacy theory of behavioral change. *Cognitive Therapy and Research*, *1*(4), 287–310. https://doi.org/10.1007/BF01663995

Banerjee, A., Breza, E., Chandrasekhar, A. G., Duflo, E., Jackson, M. O., & Kinnan, C. (2024). Changes in social network structure in response to exposure to formal credit markets. *Review of Economic Studies*, *91*(3), 1331–1372. https://doi.org/10.1093/restud/rdad065

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, *86*(6), 2453–2490. https://doi.org/10.1093/restud/rdz008

Battu, H., Belfield, C. R., & Sloane, P. J. (2003). Human capital spillovers within the workplace: Evidence for great britain*. *Oxford Bulletin of Economics and Statistics*, *65*(5), 575–594. https://doi.org/10.1111/j.1468-0084.2003.00062.x

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? *American Economic Review*, *111*(6), 1918–1943. https://doi.org/10.1257/aer.20200295

Becker, M., Kocaj, A., Jansen, M., Dumont, H., & Lüdtke, O. (2022). Class-average achievement and individual achievement development: Testing achievement composition and peer spillover effects using five german longitudinal studies. *Journal of Educational Psychology*, *114*(1), 177–197. https://doi.org/10.1037/edu0000519

Bentsen, K. H., Munch, J. R., & Schaur, G. (2019). Education spillovers within the workplace. *Economics Letters*, *175*, 57–59. https://doi.org/10.1016/j.econlet.2018.11.025

Brade, R. (2024). Short-term events, long-term friends? freshman orientation peers and academic performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4793902

Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality*. *The Quarterly Journal of Economics*, *133*(2), 611–663. https://doi.org/10.1093/qje/qjx041

Buunk, B. P., & Ybema, J. F. (2003). Feeling bad, but satisfied: The effects of upward and downward comparison upon mood and marital satisfaction. *British Journal of Social Psychology*, *42*(4), 613–628. https://doi.org/10.1348/014466603322595301

Caeyers, B., & Fafchamps, M. (2024). Exclusion bias and the estimation of peer effects. *Journal of Human Resources*, 1120–11337R2. https://doi.org/10.3368/jhr.1120-11337R2

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias*. *The Quarterly Journal of Economics*, *134*(3), 1163–1224. https://doi.org/10.1093/qje/qjz008

Carter, M., Laajaj, R., & Yang, D. (2021). Subsidies and the african green revolution: Direct effects and social network spillovers of randomized input subsidies in mozambique.

American Economic Journal: Applied Economics, *13*(2), 206–229. https://doi.org/10.1257/app.20190396

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, *329*(5996), 1194–1197. https://doi.org/10.1126/science.1185231

Chandrasekhar, A. G., Duflo, E., Kremer, M., Pugliese, J. F., Robinson, J., & Schilbach, F. (2022). Blue spoons: Sparking communication about appropriate technology use. *NBER Working Paper 30423*. https://doi.org/http://www.nber.org/papers/w30423

Chen, S., & Hu, Z. (2024). How competition shapes peer effects: Evidence from a university in china. *Review of Economics and Statistics*. https://doi.org/https://doi.org/10.1162/rest_a_01471

Cornelissen, T., Dustmann, C., & Schönberg, U. (2017). Peer effects in the workplace. *American Economic Review*, *107*(2), 425–456. https://doi.org/10.1257/aer.20141300

Daysal, N. M., Simonsen, M., Trandafir, M., & Breining, S. (2022). Spillover effects of early-life medical interventions. *The Review of Economics and Statistics*, *104*(1), 1–16. https://doi.org/10.1162/rest_a_00982

Drago, R., & Garvey, G. T. (1998). Incentives for helping on the job: Theory and evidence. *Journal of Labor Economics*, *16*(1), 1–25. https://doi.org/10.1086/209880

Duflo, E., Keniston, D., Suri, T., & Zipfel, C. (2023). Chat over coffee? diffusion of agronomic practices and market spillovers in rwanda. *National Bureau of Economic Research, No. w31368*. https://doi.org/10.3386/w31368

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals [Publisher: Annual Reviews]. *Annual Review of Psychology*, *53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., & Walker, M. (2022). General equilibrium effects of cash transfers: Experimental evidence from kenya. *Econometrica*, *90*(6), 2603–2643. https://doi.org/10.3982/ECTA17945

Fallucchi, F., Fatas, E., Kölle, F., & Weisel, O. (2021). Not all group members are created equal: Heterogeneous abilities in inter-group contests. *Experimental Economics*, *24*(2), 669–697. https://doi.org/10.1007/s10683-020-09677-5

Fiske, S. T. (2010). Envy up, scorn down: How comparison divides us. *American Psychologist*.

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, *26*(4), 331–362. https://doi.org/10.1002/job.322

Gee, L. K., Jones, J., & Burke, M. (2017). Social networks and labor markets: How strong ties relate to job finding on facebook's social network. *Journal of Labor Economics*, *35*(2), 485–518. https://doi.org/10.1086/686225

Gee, L. K., Jones, J. J., Fariss, C. J., Burke, M., & Fowler, J. H. (2017). The paradox of weak ties in 55 countries. *Journal of Economic Behavior & Organization*, *133*, 362–372. https://doi.org/10.1016/j.jebo.2016.12.004

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, *78*(6), 1360–1380.

Guryan, J., Kroft, K., & Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, *1*(4), 34–68. https://doi.org/10.1257/app.1.4.34

Hannum, E. (2005). Market transition, educational disparities, and family strategies in rural china: New evidence on gender stratification and development. *Demography*, *42*(2), 275–299. https://doi.org/10.1353/dem.2005.0014

Hannum, E., Liu, X., & Wang, F. (2021). Estimating the effects of educational system consolidation: The case of china's rural school closure initiative. *Economic Development and Cultural Change*, *70*(1), 485–528. https://doi.org/10.1086/711654

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. *NBER Working Paper 7867*. https://doi.org/10.3386/w7867

Islam, A., Vlassopoulos, M., Zenou, Y., & Zhang, X. (2021). Centrality-based spillover effects. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3880728

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85–108. https://doi.org/10.1257/app.1.4.85

Jaravel, X., Petkova, N., & Bell, A. (2018). Team-specific capital and innovation [Publisher: American Economic Association]. *The American Economic Review*, *108*(4), 1034–1073. Retrieved September 23, 2024, from https://www.jstor.org/stable/26527997

Kusumawardhani, P. N. (2022). Spillover effects of investment in early childhood education and development (ECED) centers: Evidence from indonesia. *Education Economics*, *30*(6), 590–611. https://doi.org/10.1080/09645292.2021.2019196

Lavy, V., Paserman, M. D., & Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal*, *122*(559), 208–237. https://doi.org/10.1111/j.1468-0297.2011.02463.x

Lazear, E. P. (1989). Pay equality and industrial politics [Publisher: University of Chicago Press]. *Journal of Political Economy*, *97*(3), 561–580. Retrieved September 30, 2024, from https://www.jstor.org/stable/1830455

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, *19*(3), 355–365. https://doi.org/10.1016/j.lindif.2008.10.009

Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, *50*(11), 1477–1490. https://doi.org/10.1287/mnsc.1030.0136

List, J., Momeni, F., Vlassopoulos, M., & Zenou, Y. (2023). Neighborhood spillover effects of early childhood interventions. *Centre for Economic Policy Research*.

Ma, Y., Fairlie, R., Loyalka, P., & Rozelle, S. (2024). Isolating the "tech" from EdTech: Experimental evidence on computer-assisted learning in china. *Economic Development and Cultural Change*, *72*(4), 1923–1962. https://doi.org/10.1086/726064

Malouff, J. M., Emmerton, A. J., & Schutte, N. S. (2013). The risk of a halo bias as a reason to keep students anonymous during grading [Publisher: SAGE Publications Inc]. *Teaching of Psychology*, *40*(3), 233–237. https://doi.org/10.1177/0098628313487425

Manski, C. F. (1995). *Estimation problems in the social sciences*. Harvard University Press, Cambridge, Massachusetts.

Marmaros, D., & Sacerdote, B. (2006). How do friendships form? *The Quarterly Journal of Economics*, *121*(1), 79–119.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *35*(4).

Noriega-Campero, A., Almaatouq, A., Krafft, P., Alotaibi, A., Moussaïd, M., & Pentland, A. (2018). The wisdom of the network: How adaptive networks promote collective intelligence. https://doi.org/10.48550/arXiv.1805.04766

OECD. (2014). *PISA 2012 technical report*. www.oecd.org/pisa/pisaproducts/PISA%2020%2012%20Technical%20Report_Chapter%203.pdf

Pedersen, D. E., Swenberger, J., & Moes, K. E. (2017). School spillover and college student health. *Sociological Inquiry*, *87*(3), 524–546. https://doi.org/10.1111/soin.12161

Pitsia, V., Biggart, A., & Karakolidis, A. (2017). The role of students' self-beliefs, motivation and attitudes in predicting mathematics achievement: A multilevel analysis of the programme for international student assessment data. *Learning and Individual Differences*, *55*, 163–173. https://doi.org/10.1016/j.lindif.2017.03.014

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology*, *141*, 83–100. https://doi.org/10.1016/j.jecp.2015.07.014

Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology*, *36*(1), 91–115. https://doi.org/10.1146/annurev.soc.34.040507.134743

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory : Basic psychological needs in motivation, development, and wellness*. Guilford Publications. https://doi.org/10.1521/978.14625/28806.

Simons, J., Vansteenkiste, M., Lens, W., & Lacante, M. (2004). Placing motivation and future time perspective theory in a temporal perspective. *Educational Psychology Review*, *16*(2), 121–139. https://doi.org/10.1023/B:EDPR.0000026609.94841.2f

Thien, L. M., Darmawan, I. G. N., & Ong, M. Y. (2015). Affective characteristics and mathematics performance in indonesia, malaysia, and thailand: What can PISA 2012 data tell us? *Large-scale Assessments in Education*, *3*(1), 3. https://doi.org/10.1186/s40536-015-0013-z

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of applied psychology*, *4*(1), 25–29.

Vasilaky, K. N., & Leonard, K. L. (2018). As good as the networks they keep? improving outcomes through weak ties in rural uganda. *The University of Chicago*, *66*(4). https://doi.org/https://doi.org/10.1086/697430

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*(1), 21–39. https://doi.org/10.1016/0883-0355(89)90014-1

Wu, Y. (2024). Disparities in primary and secondary education for girls in rural china. *Frontiers in Educational Research*, *7*(1). https://doi.org/10.25236/FER.2024.070125

Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: Evidence and explanations. *Personality and Social Psychology Review*, *14*(4), 368–384. https://doi.org/10.1177/1088868310366144

Appendix A. Appendix tables

Table A.1 – Randomization Schedule of the Program

	CAL	Workbook	Control
Boarding	40-minute/week after school in the computer lab	40-minute/week after school in the classroom	No interventions
4,024	1,345	1,289	1,390
nonboarding	Leaving school after class, no interventions		
6,414	2,061	2,093	2,260

Note: The table shows the randomization schedule of the program with the number of participants in each category.

Table A.2 – Impact of Two Interventions on Boarding Students' Math Performance and Class Grades

	Test Scores		Class Grades	
	(1)	(2)	(3)	(4)
CAL Program	0.033	0.032	1.743*	1.758*
	(0.039)	(0.039)	(0.919)	(0.922)
Workbook Treatment	-0.026	-0.029	1.531*	1.603*
	(0.046)	(0.046)	(0.877)	(0.876)
Additional Controls	No	Yes	No	Yes
R^2	0.432	0.436	0.300	0.308

Note: Table is created based on results from Ma et al. (2024). All columns control for baseline measures (test score or class grade). Even-numbered columns include additional covariates: liking math (scale 1-100), student age (years), gender, father graduated junior high, mother graduated junior high, teacher characteristics (teacher gender, teacher experience in years, teacher attended college) and class characteristics (number of boarding students in the class, class size). Cluster-robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.3 – Descriptive Statistics

Variable	Mean	SD	Min	Max	N
Fourth grade	0.304	0.460	0	1	6414
Fifth grade	0.350	0.477	0	1	6414
Sixth grade	0.345	0.476	0	1	6414
Gender (1=male; 0=female)	0.530	0.499	0	1	6414
Mother graduated from primary school	0.434	0.496	0	1	6414
Father graduated from primary school	0.516	0.500	0	1	6414
Family asset index	0.275	1.638	−1.874	3.561	6414
Number of family members	5.036	1.534	2	17	6414
Math test score at baseline	0.028	0.979	−4.294	1.973	6414
Has boarding peers at baseline	0.608	0.488	0	1	6414
Teacher's gender (1=male; 0=female)	0.525	0.499	0	1	340
Teacher's age (years)	37.115	9.074	22.168	60.167	340
Teacher graduated from two-year college	0.632	0.482	0	1	340
Class size	40.255	13.312	5	76	340
Number of nonboarding students	27.311	12.159	1	61	340

Note: The family asset index is derived from survey data on household items. It includes ownership of various household items, including computers, internet-connected devices, bicycles, microwaves, refrigerators, and air conditioners. We summed the items per household, with binary items counted as 1 or 0. These sums were then standardized across the sample to create the index.
Source: Authors' survey.

Table A.4 – Treatment Effect on nonboarding Student Ranking within Class

	(1) Whole Sample	(2) Has Boarding Peers	(3) No Boarding Peers
CAL Program	-0.096	-0.226	0.270
	(0.183)	(0.284)	(0.424)
Workbook Treatment	-0.092	-0.882***	1.017**
	(0.196)	(0.323)	(0.403)
Control Mean	19.693	19.586	19.872
Observations	6,414	3,898	2,516
R-squared	0.455	0.469	0.444

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level). Missing values are due to some students not answering the questions.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.5 – Treatment Effect on nonboarding Student Ranking within Group

	(1) Whole Sample	(2) Has Boarding Peers	(3) No Boarding Peers
CAL Program	-0.048	-0.021	-0.068
	(0.051)	(0.069)	(0.066)
Workbook Treatment	-0.114**	-0.134*	-0.050
	(0.051)	(0.071)	(0.059)
Control Mean	3.181	3.548	2.573
Observations	6,414	3,898	2,516
R-squared	0.394	0.359	0.425

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level).
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.6 – Treatment Effect on nonboarding Students, Excluding Those Who Changed Boarding Status

	(1) Whole Sample	(2) Has Boarding Peers	(3) No Boarding Peers
CAL Program	0.015	0.003	0.044
	(0.040)	(0.043)	(0.055)
Workbook Treatment	-0.044	-0.091**	0.024
	(0.038)	(0.043)	(0.053)
Control Mean	0.029	0.081	-0.053
Observations	5,882	3,513	2,369
R-squared	0.417	0.415	0.434

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level). 535 students (57.25% of the sample) changed from nonboarding to residing on campus during the second semester of the academic year and were excluded during analysis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.7 – How Peer Performance Affects Nonboarding Students by Peer Type

	(1) Peers are Boarding Students	(2) Peers are nonboarding Students
Peer scores (for nonboarding students)	0.083***	0.071***
	(0.020)	(0.017)
CAL group (1=yes; 0=no)	-0.000	0.010
	(0.041)	(0.042)
CAL × Score of paired student	-0.015	0.022
	(0.029)	(0.022)
Workbook group (1=yes; 0=no)	-0.096**	-0.100**
	(0.045)	(0.043)
Workbook × Score of paired student	0.067**	0.049*
	(0.031)	(0.026)
Standardized score at baseline	0.551***	0.538***
	(0.023)	(0.024)
Score of paired student at baseline	-0.009	-0.010
	(0.015)	(0.011)
Control Mean	0.100	0.0978
Observations	7,854	14,154
R-squared	0.434	0.411

Note: The dependent variable is the standardized math test score of nonboarding students at follow-up. Column (1) shows results when nonboarding students' peers are boarding students, and Column (2) shows results when nonboarding students' peers are other nonboarding students. The analysis examines how nonboarding students' performance is affected by their peers' performance within the same classroom. Only boarding students received treatment (CAL or workbook interventions). Student and teacher characteristics for both the nonboarding student and their peers are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at the class level).
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.8 – Compare Treatment Effect Between Male and Female

	(1) Math test score	(2) Instrumental motivation	(3) Math effort worthwhile	(4) Math for future job
Panel A: Has boarding peers				
CAL Program	-0.026	-0.046	-0.025	-0.019
	(0.050)	(0.053)	(0.039)	(0.043)
Workbook	-0.111**	-0.119**	-0.099***	-0.068*
	(0.050)	(0.048)	(0.036)	(0.041)
Gender (1=male; 0=female)	0.008	-0.046	-0.056*	-0.016
	(0.038)	(0.039)	(0.032)	(0.034)
CAL*Gender	0.062	0.082	0.032	0.076
	(0.058)	(0.064)	(0.046)	(0.057)
Workbook*Gender	0.047	0.121**	0.104**	0.038
	(0.059)	(0.057)	(0.046)	(0.050)
Control Mean	0.071	0.038	3.441	3.288
Observations	3,898	3,870	3,884	3,891
R-squared	0.422	0.132	0.093	0.062
Panel B: No boarding peers				
CAL Program	0.047	-0.006	-0.020	0.019
	(0.075)	(0.075)	(0.052)	(0.057)
Workbook	-0.055	-0.005	-0.023	0.034
	(0.067)	(0.060)	(0.049)	(0.050)
Gender (1=male; 0=female)	0.032	-0.043	-0.038	-0.057
	(0.060)	(0.051)	(0.045)	(0.049)
CAL*Gender	0.001	0.027	0.019	0.021
	(0.084)	(0.074)	(0.060)	(0.068)
Workbook*Gender	0.124	-0.069	-0.038	-0.039
	(0.079)	(0.073)	(0.060)	(0.066)
Control Mean	-0.063	-0.013	3.412	3.249
Observations	2,516	2,497	2,511	2,506
R-squared	0.436	0.131	0.095	0.070

Note: The model controls for outcomes at baseline. Student and teacher characteristics are controlled. All estimates include strata fixed effects. Robust standard errors are in parentheses (adjusted for clustering at class level). Missing values are due to some students not answering the questions.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix B. Appendix figures

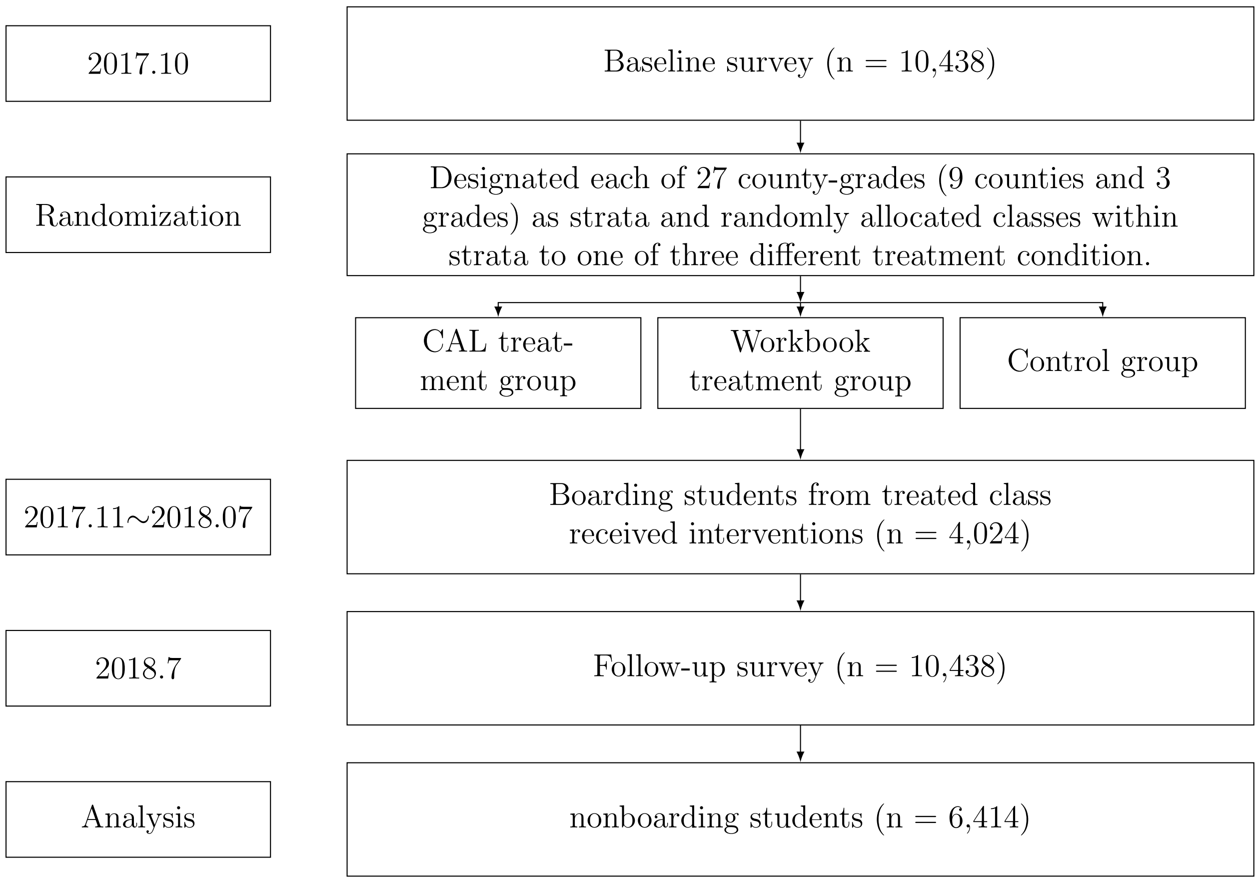


Figure B.1 – Experimental Timeline and Design

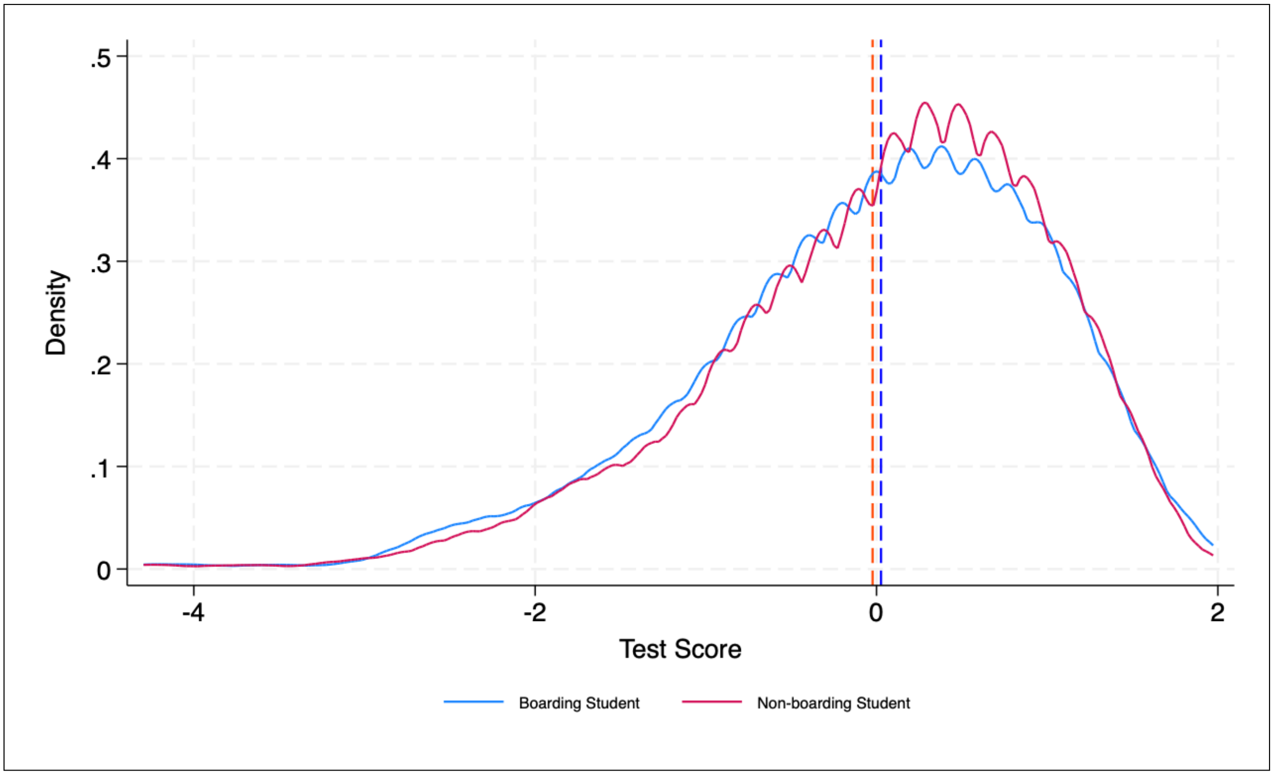


Figure B.2 – Distribution of Standardized Math Test Score at Baseline between Boarding and Nonboarding Students

Note: This figure shows the distribution of standardized math test scores at baseline for boarding and nonboarding students. The x-axis represents the standardized score, and the y-axis represents the density.

同学1 (classmates 1)	**同学2 (classmates 2)**	**同学3 (classmates 3)**	**同学4 (classmates 4)**	**同学5 (classmates 5)**
1.	**2.**	**3.**	**4.**	**5.**
同学6 (classmates 6)	**同学7 (classmates 7)**	**同学8 (classmates 8)**	**同学9 (classmates 9)**	**同学10 (classmates 10)**
6.	**7.**	**8.**	**9.**	**10.**

Figure B.3 – Study Partner List

Note: This figure displays the form used to collect study partners' names.

Class Id: **Teachers Name:**

Podium

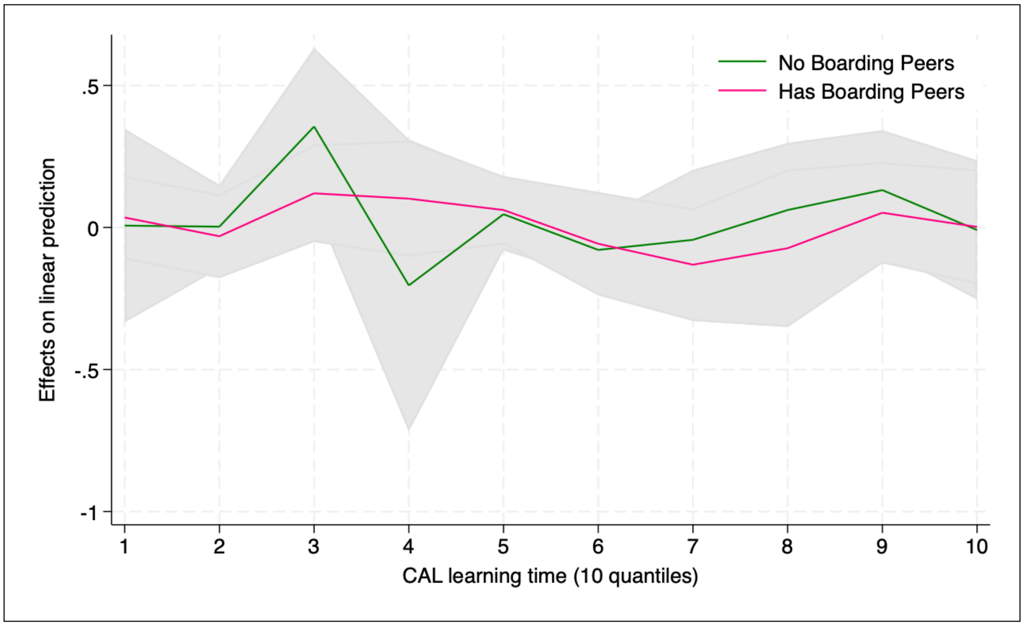
Door

1		2		3		4		5		6		7		8
Name	ID	Name	ID	Name	ID	Name	ID	Name	ID	Name	ID	Name	ID	Name

Figure B.4 – Seat Distribution Table

Note: This figure displays the form used to record seat distribution within the classroom. The first column starts from door, and the first row starts from podium.

Panel A: Average marginal effects of CAL program



Panel B: Average marginal effects of workbook program

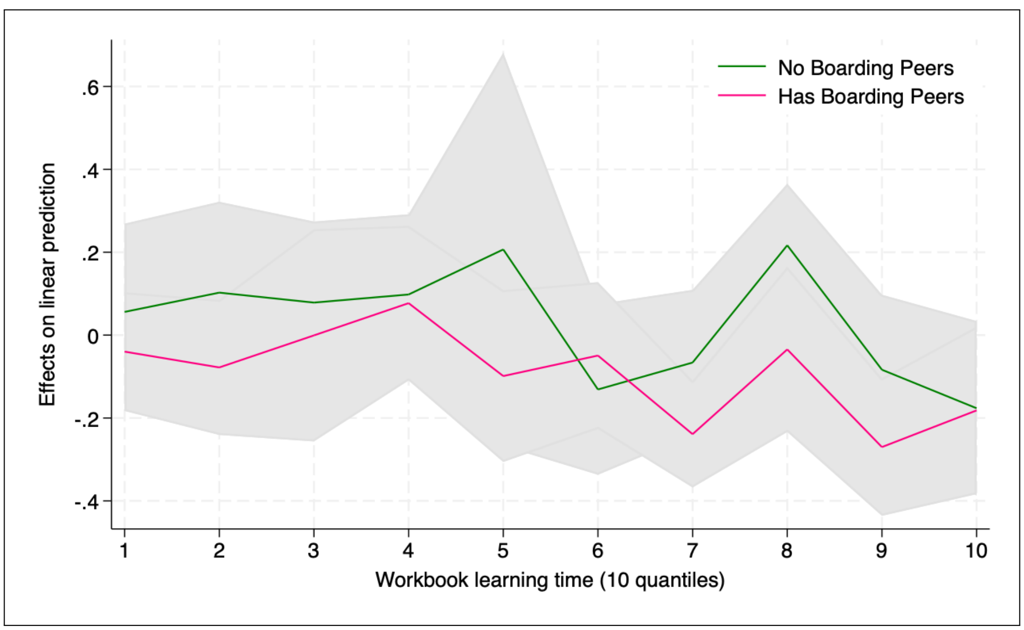


Figure B.5 – Average Marginal Effects by Treatment Intensity

Note: These figures show the average marginal effects of CAL (top) and Workbook (bottom) by treatment intensity. The x-axis represents learning time in 10 quantiles, and the y-axis shows the effects on linear prediction. Green lines represent effects for students with no boarding peers, while pink lines represent effects for students with boarding peers. The gray shaded areas indicate confidence intervals. *Source:* Authors' Survey.

Appendix C. Items of Math Learning Attitude scales

Anxiety included 5 items: a) I often worry that it will be difficult for me in mathematics classes; b) I get very tense when I have to do mathematics homework; c) I get very nervous doing mathematics problems; d) I feel helpless when doing a mathematics problem; e) I worry that I will get poor grades in mathematics.

Self-concept included 5 items: a) I am just not good at mathematics; b) I get good grades in mathematics; c) I learn mathematics quickly; d) I have always believed that mathematics is one of my best subjects; e) In my mathematics class, I understand even the most difficult work.

Intrinsic motivation included 4 items: a) I enjoy reading about mathematics; b) I look forward to my mathematics lessons; c) I do mathematics because I enjoy it; d) I am interested in the things I learn in mathematics.

Instrumental motivation included 4 items: a) Making an effort in mathematics is worth it because it will help me in the work that I want to do later on; b) Learning mathematics is worthwhile for me because it will improve my career prospects and chance; c) Mathematics is an important subject for me because I need it for what I want to study later on; d) I will learn many things in mathematics that will help me get a job.

Appendix D. Heterogeneous Treatment Effects

To examine whether treatment effects vary across students, we estimate heterogeneous treatment effects using a causal forest approach. Let Y_i denote the endline standardized test score for student i, and let $W_i \in \{\text{CAL}, \text{Workbook}\}$ indicate treatment assignment. We estimate individualized treatment effects $\tau(X_i)$ from the partially linear model

$$Y_i = \mu(X_i) + \tau(X_i)W_i + \varepsilon_i, \tag{3}$$

where X_i is a vector of pre-treatment covariates and $\mu(\cdot)$ is an unknown function capturing baseline outcome heterogeneity. The causal forest flexibly partitions the covariate space to recover $\tau(X_i)$ without imposing parametric restrictions or pre-specified subgroups.

The covariate vector X_i includes *baseline academic achievement* (baseline standardized test score), *baseline exposure to learning inputs*, and a rich set of *baseline socio-emotional and motivational skills*, including cooperation, anxiety, self-concept, intrinsic motivation, and instrumental motivation. We additionally control for baseline demographic and household characteristics and include randomization strata fixed effects. For each student, the model yields a predicted individualized treatment effect $\widehat{\tau}_i$ and an associated standard error, which we use to construct pointwise 90 percent confidence intervals.

Figure B.6 shows substantial dispersion in predicted treatment effects of the CAL intervention around the conditional average treatment effect. While the average effect is positive, the distribution reveals meaningful heterogeneity in student responses, with a nontrivial share of students exhibiting effects close to zero. Confidence intervals widen in the tails, indicating greater uncertainty for extreme predicted effects. These results suggest that CAL improves learning outcomes on average but does not affect all students uniformly.

Figure B.7 presents the corresponding distribution of individualized treatment effects for the workbook intervention. Relative to CAL, the workbook treatment exhibits greater dispersion in predicted effects, indicating stronger heterogeneity across students. This pattern suggests that the effectiveness of workbook-based instruction is more sensitive to baseline characteristics, including prior achievement and socio-emotional skills. Taken together, the evidence from both figures indicates that treatment impacts are heterogeneous and that targeting interventions based on observable student characteristics may enhance overall effectiveness relative to uniform implementation.

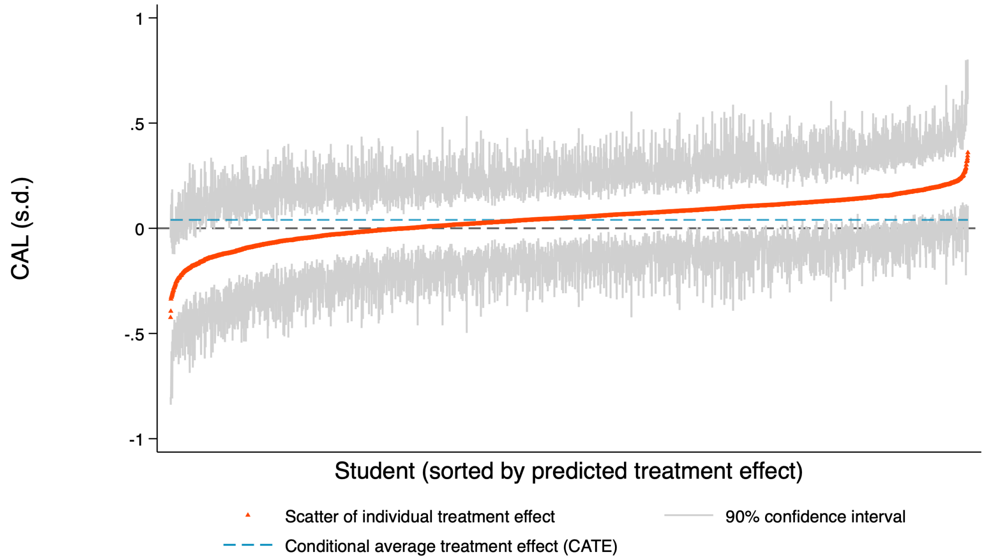


Figure B.6 – Individualized Treatment Effects of the CAL Intervention

Notes: This figure plots student-level predicted treatment effects from a causal forest using the CAL treatment indicator. Students are ordered on the horizontal axis by their predicted treatment effects (from lowest to highest). Points denote individualized predicted effects in standard deviation units, and light gray lines indicate pointwise 90% confidence intervals. The dashed horizontal line reports the conditional average treatment effect (CATE). A horizontal reference line at zero denotes no effect.

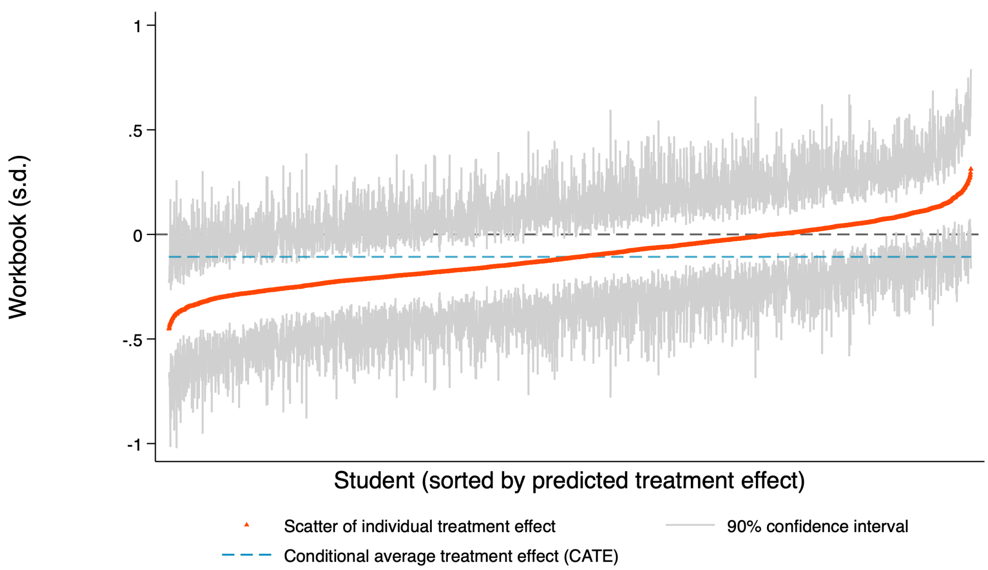


Figure B.7 – Individualized Treatment Effects of the Workbook Intervention

Notes: This figure is constructed analogously to Figure B.6, but using the workbook treatment indicator. Students are ordered by their predicted workbook treatment effects. Points denote individualized predicted effects in standard deviation units, and light gray lines indicate pointwise 90% confidence intervals. The dashed horizontal line reports the conditional average treatment effect.