Journal of Development Economics

Implicit Teacher Stereotypes Toward Left-Behind Children: A Bias-Awareness Experiment

--Manuscript Draft--

Manuscript Number:	DEVEC-D-26-00423R1
Article Type:	Registered Report Stage 1: Proposal
Section/Category:	Health, Education, gender, poverty
Keywords:	left-behind children, teacher bias, implicit association test, awareness intervention
Corresponding Author:	Christine Ho London School of Economics and Political Science SINGAPORE
First Author:	Christine Ho
Order of Authors:	Christine Ho
	Ying Liu
	Junjian Yi
Abstract:	Does bias awareness only change teacher-controlled evaluations, or can it alter broader developmental outcomes for disadvantaged children? We randomize 100 elementary school homeroom teachers in rural China to receive feedback on Implicit Association Test (IAT) scores measuring bias toward left-behind children (LBC)---students separated from migrating parents. We test whether IAT feedback affects both direct teacher behaviors---discriminatory grading and classroom interactions---and downstream student outcomes including socio-emotional outcomes and blindly graded test scores. This study provides experimental evidence on whether awareness interventions can reduce educational inequality by improving both teacher behavior and students' social outcomes.
Response to Reviewers:	

Implicit Teacher Stereotypes Toward Left-Behind Children: A Bias-Awareness Experiment

Pre-Analysis Plan May 9, 2026

Abstract

Does bias awareness only change teacher-controlled evaluations, or can it alter broader developmental outcomes for disadvantaged children? We randomize 100 elementary school homeroom teachers in rural China to receive feedback on Implicit Association Test (IAT) scores measuring bias toward left-behind children (LBC)—students separated from migrating parents. We test whether IAT feedback affects both direct teacher behaviors—discriminatory grading and classroom interactions—and downstream student outcomes including socio-emotional outcomes and blindly graded test scores. This study provides experimental evidence on whether awareness interventions can reduce educational inequality by improving both teacher behavior and students' social outcomes.

JEL Codes: D91, I21, I24, J15, J24, O15
Keywords: left-behind children, teacher bias, implicit association test, awareness intervention

1 Introduction

Teachers hold immense power over student trajectories. Beyond curriculum delivery, teachers assign grades that determine advancement, track placement, and higher education access. Mounting evidence demonstrates that teacher attitudes—particularly unconscious biases—systematically disadvantage stigmatized groups across gender (Alan et al., 2018; Carlana, 2019; Lavy and Sand, 2018; Lavy and Megalokonomou, 2024; Mengel et al., 2019; Rakshit and Sahoo, 2023), race (Dee, 2005; Papageorge et al., 2020), caste (Banerjee et al., 2025; Hanna and Linden, 2012; Ramachandran et al., 2025), ethnicity (Alan et al., 2023; Botelho et al., 2015; Burgess and Greaves, 2013), and immigration status (Carlana et al., 2022). These biases manifest in grading disparities, lowered expectations, and classroom segregation, ultimately leading to diminished academic achievement and long-term outcomes.

A critical question for education policy is whether teacher bias can be reduced through scalable interventions.[1] Alesina et al. (2024) provide a key piece of evidence: Italian middle school teachers who learned their Implicit Association Test (IAT) scores reduced grading bias against immigrant students. This finding establishes that awareness interventions can shift subjective judgments that decision-makers control directly, but leaves open whether such interventions can also produce comprehensive educational improvements for a vulnerable population whose disadvantage carries no visible markers. Do teachers not only adjust grades but also modify daily classroom interactions that build confidence and belonging? Does awareness reduce social marginalization and improve peer relationships? And do effects extend to objective achievement measures that teachers cannot directly manipulate?

We provide the first experimental evidence testing whether awareness interventions produce comprehensive educational improvements beyond teacher-controlled evaluations, studying a population for whom this question is especially pertinent: left-behind children (LBC) in rural China—students separated from parents migrating to urban cities—who display no observable markers of their family circumstances. Unlike gender or immigration status, LBC status is not obvious, meaning teacher bias may manifest as neglect rather than overt stereotyping. Beyond grading, LBC experience not only achievement gaps but also social isolation, peer rejection, and psychological distress stemming from parental absence (Fellmeth et al., 2018; Zhou et al., 2021)—outcomes that teacher bias may exacerbate and that prior experimental work has not examined. Our study is set in a developing-country context where resource constraints and heavy teacher workloads make it a non-trivial test of whether such interventions can work beyond high-income settings, and for broader development outcomes.

[1] Prior work shows that awareness interventions can shift subjective judgments more broadly, including attitudes toward female employment (Bursztyn et al., 2020) and gender bias in student evaluations of teachers (Boring and Philippe, 2021).

LBC constitute one of the most vulnerable populations in low- and middle-income country (LMIC) education systems. This phenomenon extends far beyond China: labor migration creates LBC globally, including in Southeast Asia, Latin America, Eastern Europe, and sub-Saharan Africa (UNICEF, 2025). China represents the largest concentration with approximately 70 million—or 23% of all children—left-behind in their hometowns (NBS et al., 2023), making it an ideal context to study this widespread phenomenon. Beyond the economic disadvantages common to rural students, parental absence during critical developmental periods creates additional psychological and educational challenges. LBC have lower cognitive and health outcomes (e.g., see the systematic review by Fellmeth et al. (2018) as well as Yue et al. (2020) and Zhang et al. (2014)). Moreover, negative spillovers from LBC to their classmates suggest that teacher bias could affect entire classroom environments (Huang and Zhang, 2025).

We conduct a randomized controlled trial (RCT) with 100 elementary schools in a rural county with high out-migration rates in Hunan Province. We measure teachers' implicit bias toward LBC using the IAT during a baseline survey, with one Grade 5 homeroom teacher per school. We then randomly assign teachers—stratified by median IAT scores—to receive immediate versus delayed feedback. Teachers in treatment schools receive personalized feedback via electronic message mid-semester, explaining their implicit associations and how such biases may unconsciously influence classroom behavior. Control teachers receive the same feedback until the conclusion of the study period. We track teacher behaviors (grading and classroom interactions) and comprehensive student outcomes including well-being, peer relationships, and blindly graded standardized test scores.

Our design tests a causal chain from awareness to student outcomes. We first examine whether teachers hold implicit bias against LBC at baseline and whether stronger bias correlates with worse student outcomes. We then test whether IAT feedback increases awareness and reduces teachers' explicit negative attitudes toward LBC—the core mechanism through which awareness interventions may operate. Next, we investigate whether this awareness translates into direct behavioral changes: reduced discriminatory grading and improved classroom interactions. Finally, we test whether these teacher behavior changes generate downstream improvements in student socio-emotional outcomes, ultimately improving learning outcomes measured by blindly graded standardized tests. This design allows us to identify the mechanisms through which awareness interventions work and whether benefits extend beyond subjective evaluations.

We proceed as follows. Section 2 provides contextual background on LBC in China and develops our theoretical framework. Section 3 describes the experimental design and power calculations. Section 4 details the data, outcome measures, and balance checks. Section 5 specifies the empirical strategy, robustness checks, and exploratory analyses. Section 6 presents preliminary descriptive evidence from our baseline. Section 7 describes the expected findings.

2 Contextual Framework

2.1 Background

Internal migration in China has created one of the world's largest populations of children separated from parents. China's household registration (*Hukou*) system restricts rural-to-urban migrants' access to urban public services, including education and healthcare. As a result, many rural parents migrate without their children for urban employment, leaving children under the care of grandparents or relatives to access local schools and maintain social networks (NBS et al., 2023).

Hunan Province, our study site, is among the six Chinese provinces with more than half of rural children left behind by one or both parents. Research in rural Hunan documents that over two-thirds of primary and middle school students have at least one parent working away from home (Zhang et al., 2014), reflecting substantial rural-urban migration flows characteristic of inland provinces. This high prevalence of LBC provides ideal conditions for studying teacher bias toward this vulnerable population.

Extensive research documents that LBC face significant disadvantages across multiple domains. A meta-analysis of 111 studies (264,967 children) finds detrimental effects on mental health, nutrition, and substance use (Fellmeth et al., 2018). Educational outcomes are similarly affected: the absence of both parents reduces test scores by 5.4 percentile points in math and 5.1 in Chinese in rural Hunan (Zhang et al., 2014), while early maternal migration reduces cognitive development (Yue et al., 2020). Moreover, LBC generate negative spillovers on classmates' test scores (Huang and Zhang, 2025). While existing research has examined family inputs and peer effects, teacher perceptions and differential treatment represent an unexplored mechanism through which parental absence may affect both LBC's outcomes and classroom environments more broadly.

Teachers may develop implicit associations toward left-behind children. In particular, they might hold negative stereotypes based on perceptions that LBC lack parental support or face behavioral difficulties, or conversely feel sympathy and provide compensatory attention. Regardless of direction, implicit biases may unconsciously affect grading, classroom interactions, and student outcomes. However, teachers are typically unaware of their implicit biases. Our intervention tests whether providing personalized IAT feedback can increase awareness, reduce explicit negative attitudes, and ultimately narrow educational inequalities between LBC and non-LBC students. Informal discussions with Hunan residents suggest that negative attitudes are prevalent, so we focus our theory under the assumption of negative bias and test this empirically below.

2.2 Theory and Hypotheses

2.2.1 Theory of Change

The theory of change is outlined in Figure 1. Teachers may hold implicit bias against LBC, which may manifest as reduced attention, lower expectations, or differential treatment in the classroom. The intervention delivers personalized IAT feedback to teachers, revealing their implicit associations toward left-behind students.

Mechanism. The intervention is expected to increase teachers' awareness of their own biases. We measure changes in teachers' awareness through their explicit attitudes toward LBC.

Direct Outcomes. Once aware of their bias, teachers can directly adjust their behavior in two domains. First, teachers can correct for bias when grading assignments, reducing discriminatory evaluation and narrowing the gap between LBC and non-LBC students in teacher-assigned grades.[2] Second, teachers may modify their daily classroom interactions, providing more attention and encouragement, offering more positive feedback, and demonstrating greater sensitivity to LBC students' emotional needs. These improvements in teaching engagement and teacher-student interactions are measured through students' perceptions of teacher behaviors.

Downstream Outcomes. When LBC experience more supportive and equitable treatment from teachers, multiple dimensions of their non-academic outcomes may improve. They may exhibit fewer behavioral problems, demonstrate enhanced self-confidence and higher self-expectations, and report a stronger sense of belonging at school. Beyond these individual-level changes, improvements in teacher behavior may alter the broader classroom climate by signaling more inclusive social norms to the entire peer group. This shift in classroom atmosphere may reduce the social marginalization of LBC, helping them develop better peer relationships and engage in fewer negative classroom behaviors. Together, fairer grading, improved teacher-student interactions, and enhancements in student well-being and behaviors may reinforce LBC's academic engagement and learning effort, ultimately translating into gains in blindly graded standardized test scores.

2.2.2 Hypotheses: Bias and Intervention Effects

Our theory of change gives rise to several hypotheses. We formalize the theory in a stylized model in Appendix A, which shows how teacher bias creates achievement gaps and how awareness interventions can narrow these gaps through direct and downstream channels.

[2] We complement this measure with a vignette-based picture grading task that provides an additional measure of grading bias.

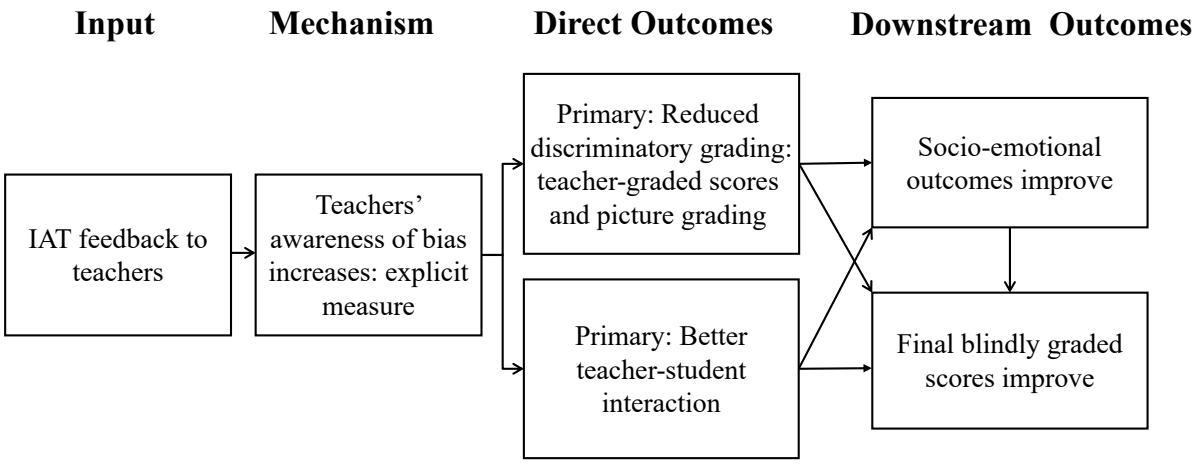


Figure 1: Theory of Change

Baseline Hypothesis: Baseline Outcomes

Hypothesis 1 (Baseline Bias)**.** *(a) Teachers exhibit implicit bias against left-behind children (positive IAT D-scores). (b) Stronger teacher implicit bias is associated with larger gaps between LBC and non-LBC in baseline academic and non-academic outcomes.*

Mechanism Hypothesis: Awareness

Hypothesis 2 (Awareness and Explicit Attitudes)**.** *Receiving personalized IAT feedback increases teachers' awareness of their bias, reducing explicit negative attitudes toward LBC.*

Primary Hypotheses: Direct Outcomes

Hypothesis 3 (Direct Effect on Grading)**.** *The intervention reduces discriminatory grading, narrowing the achievement gap between LBC and non-LBC in (a) teacher-graded scores on the standardized assignment administered immediately after the intervention and (b) a vignette-based picture grading task.*

Hypothesis 4 (Direct Effect on Teacher-Student Interactions)**.** *The intervention improves teacher-student interactions, reducing the gap in how LBC and non-LBC perceive their interactions with teachers.*

Secondary Hypotheses: Downstream Outcomes

Hypothesis 5 (Downstream Effect on Socio-emotional Outcomes)**.** *Through improved teacher-student interactions, the intervention enhances non-academic outcomes for LBC relative to non-LBC, including reduced behavioral problems, improved well-being, better peer relationships, and fewer negative classroom behaviors.*

Hypothesis 6 (Downstream Effect on Learning Outcomes)**.** *Through improved teacher-student interactions and enhanced student well-being, the intervention narrows the achievement gap between LBC and non-LBC in blindly graded standardized test scores.*

3 Experimental Design

3.1 The Implicit Association Test

To measure teachers' implicit bias toward LBC, we employ the IAT, a widely used psychological measure developed by social psychologists (Greenwald et al., 1998). The test is designed to capture automatic associations that operate outside conscious awareness. It compares response times across sorting tasks that pair target categories (LBC versus non-LBC) with evaluative attributes (positive versus negative traits). Following established protocols, faster response times when LBC are paired with negative attributes, relative to positive attributes, are captured by positive standardized D-scores, indicating implicit bias against LBC. Following standard interpretation guidelines, D-scores between -0.15 and 0.15 indicate negligible bias; between 0.15 and 0.35 in absolute value indicate slight bias; between 0.35 and 0.60 in absolute value indicate moderate bias; and above 0.60 in absolute value indicate strong bias (Greenwald et al., 2009). The LBC and non-LBC stimuli were piloted with a separate sample of 20 teachers and students prior to the study, and words consistently flagged as ambiguous were dropped or revised. Detailed procedures for IAT administration are provided in Appendix B.1.

Recent work in economics has increasingly adopted the IAT to study implicit bias and its consequences. Prior studies document that IAT scores predict labor market outcomes such as call-back rates (Rooth, 2010) and on-the-job performance (Glover et al., 2017). In educational settings, implicit bias measured by the IAT has been shown to affect teachers' track recommendations (Carlana et al., 2022) and contribute to gender gaps in academic performance (Carlana, 2019). Experimental evidence further suggests that IAT-measured bias predicts biased beliefs and suboptimal belief updating even in the presence of objective performance information (Reuben et al., 2014).

We acknowledge that the IAT has well-documented limitations. The predictive validity of the IAT remains debated, with several meta-analyses reporting that IAT scores correlate only weakly with discriminatory behavior (Blanton et al., 2009; Oswald et al., 2013). However, much of this

critical evidence comes from small-sample laboratory studies that lack information on actual behavior toward stigmatized groups outside the lab (Carlana, 2019). Recent field-based studies in economics have, by contrast, documented significant correlations between IAT scores and real-world outcomes including hiring, on-the-job performance, and teachers' track recommendations (Carlana, 2019; Carlana et al., 2022; Glover et al., 2017; Reuben et al., 2014; Rooth, 2010). We acknowledge these debates directly in the feedback provided to teachers, where we explain that the IAT measures implicit attitudes rather than behaviors and that such attitudes derive from cultural and social context.

Despite these limitations, the IAT offers several advantages that make it well suited for the present study. First, it captures automatic associations that are difficult to elicit through self-reported measures. Teachers may be unwilling or unable to explicitly report differential attitudes toward LBC, or may not be consciously aware of such attitudes. Second, unlike explicit attitude measures, IAT scores are difficult to strategically manipulate. Because the test relies on reaction times under time pressure, respondents have limited scope for consciously adjusting their responses. This concern is further alleviated in our context: the teachers in our sample are predominantly rural teachers in China, for whom the IAT is unfamiliar and not commonly used in professional or research settings. As a result, the IAT provides a measure of implicit attitudes that is unlikely to be systematically distorted by strategic response behavior. We complement the IAT with explicit attitude measures in our teacher surveys.

3.2 The Field Experiment

Research Context. This study takes place in a relatively poor rural county in Hunan Province. Based on local government records, the county has a population of approximately 933,000, of whom about 222,000 are migrant workers employed outside the county. In 2023, per capita disposable income in this county was RMB 21,718, substantially below the provincial average of RMB 35,895 in Hunan (Hunan Provincial Bureau of Statistics, 2024) and the national average of RMB 39,218 (National Bureau of Statistics of China, 2024). In this setting, a substantial share of primary school students (27%) live apart from one or both parents and are cared for by grandparents or other relatives. This institutional and demographic context provides a particularly relevant environment for studying teachers' implicit attitudes toward LBC and their potential consequences for educational outcomes.

The Homeroom Teacher System. In Chinese elementary schools, each class is assigned a homeroom teacher (*Banzhuren*) who serves as the primary point of contact between the school and students' families. Unlike subject teachers who only interact with students during specific lessons,

the homeroom teacher plays a multifaceted role in students' school lives. Specifically, homeroom teachers are responsible for: (1) monitoring students' academic progress across all subjects and coordinating with subject teachers; (2) managing classroom discipline and addressing behavioral issues; (3) organizing class activities and fostering classroom culture; (4) maintaining regular communication with parents or guardians about students' overall development; and (5) providing guidance on students' social-emotional development and peer relationships.

In the local context, the homeroom teacher typically teaches one core subject (usually Chinese) to their assigned class and spends significantly more time with students than other teachers—including supervising morning readings, lunch breaks, and after-school activities. Importantly, homeroom teachers are well-positioned to identify LBC in their class, as the left-behind status of students is officially registered and documented. Moreover, the same homeroom teacher usually follows the students across terms in primary school. Combined with their responsibility of maintaining regular communication with parents or guardians and monitoring students' overall development, homeroom teachers are well aware of individual students' family circumstances.

Recruitment and Sampling. We employ a two-stage sampling design with assistance from the local Bureau of Education. First, we randomize the order of towns to visit within the county. Second, within each town, we attempt to recruit all public primary schools with Grade 5 classes until we reach our target of 100 schools. This ensures geographic representation while maintaining logistical feasibility. Recruitment is conducted through phone calls and in-person visits by the research team. We track and report: (1) the number of towns contacted, (2) the number of schools contacted within each town, (3) participation rates at both town and school levels, and (4) reasons for non-participation (if any). These will be reported together with basic statistics comparing our sample schools to schools in rural China (if available) once recruitment is completed.

The recruitment process begins with contacting school principals. During the initial contact, the research team introduces the purpose of the study, outlines the research procedures, and explains the requirements for school participation. Upon obtaining the principal's approval, the research team meets with the relevant Grade 5 teachers to further explain the study objectives, data collection process, and the voluntary nature of participation. Teachers' agreement to participate is obtained prior to any further recruitment activities. After teachers have agreed to participate, they assist the research team by distributing recruitment materials and informed consent link to parents or guardians through existing class WeChat groups, which are the standard communication channel between schools and families in the study context. Parents receive detailed study information and provide consent independently. For students, the research team conducts a verbal briefing to explain the study purpose, procedures, and their rights as participants, including the voluntary nature of participation and the right to withdraw at any time. Student assent is obtained following

this briefing. Student participation in the study requires both parental consent and student assent.

We focus on Grade 5 students, a group old enough to understand and respond reliably to study questionnaires and tasks, yet young enough that parental absence may have meaningful cognitive and socio-emotional consequences (Heckman, 2006). Younger students may lack the cognitive capacity to complete the questionnaires accurately, while Grade 6 students face a more intense academic schedule as they prepare to transition to middle school. In most rural schools in our study context, there is only one Grade 5 class; in schools with multiple Grade 5 classes, one class is randomly selected to take part in the study.

In total, the study aims to recruit approximately 100 schools, involving around 100 school principals and 100 Grade 5 homeroom teachers. The student-level sample consists of approximately 2,700 Grade 5 students, although the number of students may vary by school due to differing class sizes across schools.

Randomization. The study adopts a school-level randomized controlled design. After completion of baseline surveys and administration of the IAT to teachers, participating schools are randomly assigned to treatment or control groups in a 1:1 ratio (50 schools per group), stratified by whether the participating teacher's baseline IAT score is above or below the median.

Intervention. Teachers in treatment schools receive individualized feedback on their IAT results in early May 2026. This intervention is designed to be light-touch, cost-effective, and scalable. The feedback is delivered online: teachers receive a personalized report via WeChat. Following best practices from Alesina et al. (2024), the feedback explains what the IAT measures, how to interpret individual IAT scores, and how implicit associations may influence behaviors unconsciously. This feedback is accompanied by general information designed to raise awareness of implicit bias and its potential consequences for student evaluation and classroom interactions. Teachers in control schools complete the IAT but do not receive any feedback until the conclusion of the study period. Details of the intervention content (feedback template) are provided in Appendix B.2.

3.3 Power Calculation

We conduct power calculations to determine the minimum detectable effect (MDE) for our outcomes of interest. Given that randomization is at the school level, we follow standard formulas for cluster-randomized trials and account for both intracluster correlation and unequal cluster sizes. The MDE for an outcome compared across treatment and control is:

$$\text{MDE} = (z_{1-\kappa} + z_{\alpha/2}) \sqrt{\frac{1}{P(1-P)} \times \frac{\sigma^2}{N} \times (1 + (\bar{m}(1+CV^2) - 1) \times \text{ICC})} \tag{1}$$

where $z_{1-\kappa}$ and $z_{\alpha/2}$ are the standard normal critical values for power $1-\kappa$ and significance level α (two-sided test), P is the proportion of schools assigned to treatment, σ^2 is the outcome variance, N is the total sample size, $\bar{m}$ is the average cluster size, CV is the coefficient of variation of cluster sizes, and ICC is the intracluster correlation coefficient. When baseline data are available, we adjust the MDE by multiplying by $\sqrt{1-\rho^2}$, where ρ is the baseline–endline correlation of the outcome. We apply Equation (1) to teacher-level outcomes and to outcomes within the LBC subsample.

A central focus of our study is whether the intervention narrows outcome gaps between LBC and non-LBC students. This is captured by the interaction between treatment assignment and LBC status, identified by their joint variation in the data. The MDE for this interaction therefore depends not only on overall sample size but also on the share of LBC students s, with the denominator scaled by $P(1-P)\cdot s(1-s)$:

$$\text{MDE}_{\text{interaction}} = (z_{1-\kappa}+z_{\alpha/2})\sqrt{\frac{1}{P(1-P)\cdot s(1-s)}\times\frac{\sigma^2}{N}\times(1+(\bar{m}(1+CV^2)-1)\times\text{ICC})} \quad (2)$$

Table A2 reports MDEs at three levels of analysis, each tied to specific hypotheses. The average cluster size, coefficient of variation, ICCs, and baseline–endline correlations for academic outcomes are obtained from local education bureau records. The proportion of LBC ($s = 0.22$) and parameters for socio-emotional outcomes are from our partially digitized student baseline (21 schools). Baseline–endline correlations for non-academic outcomes are taken from the literature (see Table A2 for sources).

Teacher-level outcomes (Panel A). Hypotheses 2 (explicit attitudes) and the teacher-level component of Hypothesis 3 (vignette-based picture grading) are tested at the teacher level ($N = 100$). Applying Equation (1), the MDE with baseline controls is 0.43 SD at 80% power. As a benchmark, Alesina et al. (2024) find effects of around 0.6 SD for high-IAT teachers in their online experiment, in which teachers grade fictitious tests—a design structurally similar to our vignette-based picture grading task. We acknowledge that smaller effects may not be detectable at $N = 100$ and therefore frame teacher-level outcomes as complementary mechanism evidence; our primary tests of behavioral change are at the student level where power is substantially higher.

Differential treatment effect by LBC status (Panel B). Our primary parameter of interest is β_3 in Equation (6), capturing whether the intervention narrows the LBC–non-LBC gap. Applying Equation (2), MDEs with baseline controls range from 0.22 to 0.27 SD across primary outcomes (teacher-graded assignment scores and perceived teacher–student interactions under Hy-

potheses 3 and 4) and from 0.24 to 0.30 SD across secondary outcomes (socio-emotional outcomes and blindly graded test scores under Hypotheses 5 and 6), at 80% power. Our design is therefore powered to detect effects of the magnitude documented in comparable studies: Alesina et al. (2024) find a 0.27 SD reduction in discriminatory grading from personalized IAT feedback, and Boring and Philippe (2021) estimate a 0.30 SD reduction in gender bias in evaluations, both from student-level analyses.

LBC subsample (Panel C). As a complementary analysis to the gap-narrowing tests in Panel B, we report MDEs for treatment effects estimated within the LBC subsample only (estimated at $N = 594$, that is, 22% of 2,700 surveyed students), using Equation (1). This tests whether LBC outcomes improve in absolute terms in treatment schools, rather than relative to non-LBC. With baseline controls, MDEs range from 0.13 to 0.20 SD at 80% power—smaller than Panel B, reflecting the more focused comparison within the population of primary interest.

3.4 Timeline

The academic calendar in rural Hunan consists of two terms per academic year. The first term runs from September to January, and the second term runs from March to July. We outline the expected timeline for the study below. The study takes place during the second term, although we collect administrative records from the first term for baseline measures.

- **Baseline data collection (Jan–Apr 2026):** Recruitment of participating schools, teachers, and students. Collection of baseline surveys from teachers (including IAT), students, and school principals. Collection of baseline administrative data, including centrally and blindly graded final examinations from the first academic term.
- **Intervention (Early May 2026):** Random assignment of schools to treatment and control groups, stratified by teachers' baseline IAT scores (above vs. below median). Treatment teachers receive personalized feedback on implicit bias toward LBC via electronic message during the semester. Control teachers receive the same feedback at the end of the study.
- **Endline data collection (Jun–Jul 2026):** Collection of endline surveys from teachers, students, and school principals; teacher-graded assignment scores; centrally and blindly graded final examination scores from the second academic term.

4 Data

4.1 Data Collection

Data for this study are collected from multiple sources. Specifically, we collect: (i) school-based surveys from school principals, Grade 5 homeroom teachers, and students and (ii) administrative data from schools on students' academic performance, including both centrally and blindly graded examinations as well as teacher-graded assignments. Data collection takes place at two points in time. Baseline data are collected between January and April 2026, prior to treatment assignment. Endline data are collected at the end of the academic year, June-July 2026, following the intervention.

Headmaster survey. The headmaster survey includes questions on basic demographics (e.g., gender, education, teaching experience, childhood family migration background), It also gathers school-level information, such as the number of students and teachers and the proportion of left-behind students in the school. In addition, the survey includes questions on teacher evaluation and promotion practices, salary composition, and headmasters' attitudes toward LBC.

Teacher survey. The teacher survey is administered to Grade 5 homeroom teachers and consists of two main components. The first component is the IAT, which measures teachers' implicit associations toward LBC. The second component is a questionnaire that collects information similar to the headmaster survey, including teachers' demographic characteristics, class-level information, and attitudes toward LBC. In addition, the teacher survey includes measures of teachers' working hours, mental health, teaching engagement, and personality traits.

Student Surveys. The student survey collects information on students' personal characteristics and family background. It also measures students' perceptions of teacher engagement in the classroom. In addition, the survey includes measures of non-academic outcomes including behavioral index, peer relationships, self-confidence, self-expectations, negative classroom behaviors, and school sense of belonging.

Administrative Data. We collect administrative data from schools on students' academic performance including baseline test scores from centrally and blindly graded final examinations at the end of the first term. Endline academic data include both teacher-graded assignments and centrally and blindly graded final examinations at the end of the academic year.

Data management. Survey data from school principals and teachers are collected using Credamo, a secure online survey platform commonly used in China. Student surveys are administered in person using paper-based questionnaires during school hours. Personally identifiable information will be removed from survey and administrative data and replaced with anonymized participant identifiers. Identifying information will be stored separately from research data in password-protected files accessible only to the research team. All analyses will be conducted using anonymized data, and results will be reported in aggregate form.

4.2 Measures

Teacher-related measures. First, we collect teachers' background characteristics, including demographic information and teaching experience. Other key measures include:

- **Implicit attitudes.** Teachers' implicit attitudes toward LBC are measured using the IAT. The baseline IAT score forms the basis for the personalized feedback delivered during the intervention. At endline, treatment teachers are asked to identify the bias category their IAT score fell into ("no bias / slight / moderate / strong"). This serves as an objective measure of engagement with the feedback.

- **Explicit attitudes.** To measure teachers' explicit attitudes toward LBC, we present teachers with hypothetical classroom scenarios that vary in the proportion of left-behind students. For each scenario, teachers rate the classroom on three dimensions using a 5-point scale: teaching effectiveness, classroom discipline, and peer relationships among students. This approach follows the spirit of measures used in well-validated surveys examining attitudes toward migrant students (e.g., the China Education Panel Survey). These measures capture teachers' conscious beliefs and expectations about LBC.

- **Grading behaviors.** Grading-related behaviors are measured using two approaches. First, we collect teacher-assigned scores on assignments given immediately after the intervention as part of final examination preparation and before any learning effects from the intervention can fully materialize. Second, we employ a picture grading task in which teachers grade AI-generated children's drawings that implicitly signal LBC versus non-LBC status through depicted life scenarios—for example, living with grandparents (LBC) versus living with parents (non-LBC)—without being told the study's focus on LBC (see Appendix B.4). This task is administered at both baseline and endline enabling a difference-in-differences analyses.

- **Perceived teaching behaviors.** Students' perceptions of teacher behaviors are measured using student surveys. Students are asked how frequently specific teacher behaviors occur

in their classroom. For example, whether the teacher notices students' emotional states, praises students for good performance, or engages in negative interactions such as yelling. Responses are recorded on a Likert scale ranging from "never" to "always." These measures capture students' perceived day-to-day interactions with teachers and follow the approach used in Alan et al. (2024).

Student-related measures. Student-related measures will be collected using student surveys and administrative data at both baseline and endline.

- **Behavioral problems.** We measure students' behavioral problems using a 14-item behavior problem index adapted from the China Family Panel Studies (CFPS), which captures students' emotional and behavioral adjustment in the school context, including items related to conduct problems, emotional difficulties, and peer relationship problems. Higher values indicate more severe behavioral problems.

- **Self-confidence and self-expectations.** Students' self-confidence and self-expectations are measured using survey items that assess students' beliefs about their own academic abilities, confidence about the future, and future academic prospects.

- **School sense of belonging.** We measure students' school sense of belonging using survey items that assess their attendance motivation, peer acceptance, and feelings of loneliness at school. Students respond on a 4-point Likert scale. These measures capture students' psychological connection to the school environment.

- **Peer relationships.** We measure peer relationships using two approaches. First, students self-report the number of friends they have in the classroom. Second, students nominate up to five classmates whom they consider to be their close friends. From these data, we construct multiple measures of peer relationships. Social integration measures include the number of self-reported friends, out-degree (number of nominations made), and in-degree (number of nominations received). Social isolation measures include binary indicators for students who make no friendship nominations (out-degree isolation) or receive no nominations from classmates (in-degree isolation). These measures will be normalized by class size or number of responding peers in each class. Due to low absence rates in China, we expect missing responses to be low in our setting.

- **Negative classroom behaviors.** We assess negative classroom behaviors by asking students about their experiences with and participation in verbal threats, physical aggression, spreading false information, social exclusion, and online harassment. While inspired by measures

in Cunha et al. (2023), we refine and reframe all items to use age-appropriate language and to reflect a broader set of negative peer interactions in the classroom. These items capture students' involvement in such behaviors both as perpetrators and as targets.

- **Academic outcomes.** Besides teacher-graded assignments, student academic outcomes are measured using centrally and blindly graded final examinations. Baseline academic performance is measured using final examinations from the first academic term, while endline academic performance is measured using final examinations from the second academic term. Because these examinations are graded blindly by external evaluators, they provide objective measures of student achievement that are not directly influenced by teachers' grading discretion and therefore capture changes in underlying academic performance over time.

Table 1 summarizes our key variables, direct outcomes, downstream outcomes, and mechanisms, along with their data sources and measurement scales. Detailed descriptions of specific survey questions and measurement instruments are provided in the Appendix B.4.

4.3 Balance Check

We assess baseline balance using two complementary approaches. First, we report means of pre-treatment student, teacher, and school characteristics across treatment and control groups. To test for differences while reflecting our stratified randomization, we estimate strata-adjusted mean differences by regressing each variable on the treatment indicator and strata fixed effects (above vs. below median baseline IAT score), reflecting our stratified randomization design (Bruhn and McKenzie, 2009).

Second, we conduct an omnibus test of joint orthogonality by regressing treatment status on all baseline covariates (including strata fixed effects) and testing the null hypothesis that all coefficients equal zero. Because standard F-tests of joint orthogonality have been shown to over-reject the null of balance, we report p-values based on randomization inference following Kerwin et al. (2024), which are valid in finite samples and respect our stratified randomization design. These will be reported once baseline data digitization is completed.

5 Empirical Strategy

Our analyses are conducted at the teacher (school) and student levels, based on the relevant outcomes. We employ linear models in our main analyses and perform sensitivity analyses using nonlinear models for relevant outcomes.

Table 1: Snapshot of Key Variables, Outcomes, and Mechanisms

Variable	Hyp.	Source of measurement	Measure	Analytical variable
Baseline Variable				
Teachers' implicit attitudes	H1	**Teacher survey**: Implicit Association Test (IAT)	D-score (continuous)	D-score (continuous)
Mechanisms: Awareness				
Teachers' explicit attitudes	H2	**Teacher survey**: Hypothetical classrooms with varying LBC proportions	Likert scale (5-point)	Explicit bias index (3 dimensions)
Primary Outcomes: Direct				
Teacher-graded assignment scores	H3	**Administrative records**: teacher-assigned scores right after the intervention	0–100 scale	Standardized score
Picture grading task	H3	**Teacher survey**: vignette-based picture grading	0–10 scale	Grading gap (non-LBC minus LBC, standardized)
Perceived teacher behaviors	H4	**Student survey**: Perceptions of teacher classroom behaviors	Likert scale (4-point)	Teacher behavior index (3 items)
Secondary Outcomes: Downstream				
Behavioral and well-being outcomes	H5	**Student survey**: 14-item behavioral problem index, perceived academic standing, educational aspirations, confidence about the future, sense of belonging	Likert scales; Ranked categorical	Behavioral problems index (14 items); well-being index (self-perception, 3 items); well-being index (school belonging, 3 items)
Peer relationships and negative classroom behaviors	H5	**Student survey**: Self-reported friends, peer nominations, involvement in negative peer interactions	Cardinal; Network; Dummy (0/1); Frequency scale	Peer integration index (3 items); peer isolation index (2 items); perpetration index (5 items); experience index (5 items)
Blindly graded test scores	H6	**Administrative records**: centrally blindly graded final exam scores	0–100 scale	Standardized score

Note: All indices are constructed as standardized averages. See Appendix B.4 for survey questions and Appendix B.5 for index construction details.

Baseline correlational analyses We first examine whether teachers' implicit bias is associated with teacher outcomes such as their explicit attitudes toward LBC at baseline. We estimate:

$$Y_s^0 = a_0 + a_1 \text{IAT}_s + \mathbf{T_s'a_T} + \mathbf{Z_s'a_Z} + e_s, \tag{3}$$

where Y_s^0 is the baseline measure of teacher outcome in school s and IAT_s denotes the teacher's baseline IAT score. The coefficient a_1 captures the correlation between teachers' implicit bias and teachers' outcomes. $\mathbf{T}_s$ is a vector of pre-specified baseline teacher controls, including gender, age, an indicator for the teacher's own left-behind experience in childhood, and an indicator for randomization strata (above vs. below median baseline IAT score). $\mathbf{Z}_s$ is pre-specified baseline school control, including the proportion of LBC students in the school. e_s is an error term and standard errors are robust.

We then document descriptive relationships between teachers' implicit attitudes toward LBC and student outcomes using baseline data. Specifically, we examine how teachers' IAT scores are correlated with students' academic and socio-emotional baseline differences between LBC and non-LBC. To characterize the baseline relationship between teachers' implicit attitudes and student outcomes, we estimate:

$$Y_{is}^0 = \alpha_0 + \alpha_1 \,\text{LBC}_i + \alpha_2 \,(\text{IAT}_s \times \text{LBC}_i) + \mathbf{X_i'}\alpha_\mathbf{X} + \gamma_s + \varepsilon_{is}, \tag{4}$$

where Y_{is}^0 is the baseline outcome of student i in school s, IAT_s denotes the baseline IAT score of the Grade 5 homeroom teacher in school s, and LBC_i is an indicator equal to one if student i is a left-behind child (both parents absent) and zero otherwise. α_2 captures whether the baseline gap between left-behind and non-left-behind students varies systematically with teachers' IAT scores. $\mathbf{X}_i$ is a vector of pre-specified baseline student controls, including gender, only-child status, and categorical indicators for mother's education and father's education (completed compulsory education, high school or above, unknown). γ_s represents school fixed effects. Since each school contributes one Grade 5 class with one homeroom teacher, the school fixed effects absorb all school-level and teacher-level characteristics, including the teacher's IAT score and other student-invariant attributes. ε_{is} is an error term, and standard errors are clustered at the school level.

Intention-to-treat framework We estimate the causal effects of the intervention using an intention-to-treat (ITT) framework. We first examine whether the intervention affects teachers' outcomes such as explicit attitudes. At the teacher level, we estimate:

$$Y_s^1 = b_0 + b_1 \text{Feedback}_s + \mathbf{T_s'b_T} + \mathbf{Z_s'b_Z} + u_s, \tag{5}$$

where Y_s^1 denotes the endline teacher-level outcome for the teacher in school s. The treatment indicator Feedback_s equals 1 if school s is assigned to treatment (teachers receive IAT feedback) and 0 otherwise. The coefficient b_1 captures the treatment effect on teacher outcomes. $\mathbf{T}_s$ and $\mathbf{Z}_s$ are the same teacher and school controls as in Equation (3). u_s is an error term, and standard errors are robust.

When examining the effects for students' outcomes, we test whether the intervention reduces the gap in academic and socio-emotional outcomes between LBC and non-LBC. To estimate the causal effects of the intervention, we employ the following specification:

$$Y_{is}^1 = \beta_0 + \beta_1 \,\text{Feedback}_s + \beta_2 \,\text{LBC}_i + \beta_3 \,(\text{Feedback}_s \times \text{LBC}_i) + \mathbf{X_i'}\beta_{\mathbf{X}} + \mathbf{T_s'}\beta_{\mathbf{T}} + \mathbf{Z_s'}\beta_{\mathbf{Z}} + \upsilon_{is}. \quad (6)$$

Y_{is}^1 denotes the endline outcome for student i in school s. The treatment indicator Feedback_s equals 1 if school s is assigned to treatment (teachers receive IAT feedback) and 0 otherwise. $\mathbf{X}_i$ is a vector of pre-specified baseline student controls as shown in Equation (4). $\mathbf{T}_s$ and $\mathbf{Z}_s$ are the same teacher and school controls as in Equation (3). υ_{is} is an error term, and standard errors are clustered at the school level.

β_3 is our primary parameter of interest, capturing the differential treatment effect on LBC students. It measures how the intervention narrows the gap between LBC and non-LBC. β_1 and $\beta_1 + \beta_3$ capture the marginal effects of the intervention on, respectively, non-LBC and LBC. While β_3 tests whether the intervention specifically narrows the LBC/non-LBC gap, $\beta_1 + \beta_3$ captures the policy-relevant total effect on LBC students—that is, whether LBC in treatment schools fare better than LBC in control schools.

ANCOVA specification. For outcome variables measured at both baseline and endline, we include the baseline level of the dependent variable as a covariate in Eq. (5)-(6). This ANCOVA (analysis of covariance) approach improves statistical power by reducing residual variance when baseline and endline outcomes are correlated. While randomization ensures treatment and control groups are balanced on average, including baseline controls increases precision by accounting for individual-level variation in initial conditions.

Multiple hypothesis testing. To address concerns about multiple comparisons, we report sharpened q-values using the Benjamini et al. (2006) procedure to control the false discovery rate. For outcomes related to the main hypotheses—Hypothesis 2 (mechanism), 3 (grading), and 4 (interactions)—we pre-specify the following family groups:

- Direct outcomes at the teacher-level: explicit attitudes and vignette-based picture score.

- Direct outcomes at the student-level: teacher-graded scores on the standardized assessment administered immediately after the intervention and teacher-student interactions.

For downstream outcomes under Hypothesis 5 (well-being and peer relationships) and Hypothesis 6 (learning), we pre-specify the following family groups:

- Behavioral and well-being outcomes: behavioral problems index, self-perception index, school belonging index—3 outcomes.
- Peer relationships: peer integration index, peer isolation index—2 outcomes.
- Academic performance: blindly graded standardized test scores—single outcome so we will report conventional p-values.

5.1 Robustness Checks

Differential attrition analysis. If a homeroom teacher leaves between baseline and endline, the corresponding school will be excluded from the main ITT analysis. We assess differential teacher attrition by estimating the following teacher-level regression:

$$\text{TA}_s = \lambda_0 + \lambda_1 \text{Feedback}_s + \mathbf{T}_\mathbf{s}'\lambda_\mathbf{T} + \mathbf{Z}_\mathbf{s}'\lambda_\mathbf{Z} + \zeta_s, \tag{7}$$

where TA_s equals one if the homeroom teacher in school s is present at baseline but missing at endline. We test whether λ_1 differs significantly from zero, indicating differential teacher attrition across treatment arms.

We assess whether student attrition between baseline and endline differs systematically across treatment arms by estimating:

$$\text{SA}_{is} = \gamma_0 + \gamma_1 \text{Feedback}_s + \gamma_2 \text{LBC}_i + \gamma_3 (\text{Feedback}_s \times \text{LBC}_i) + \mathbf{X}_\mathbf{i}'\gamma_\mathbf{X} + \mathbf{T}_\mathbf{s}'\gamma_\mathbf{T} + \mathbf{Z}_\mathbf{s}'\gamma_\mathbf{Z} + \xi_{is}, \tag{8}$$

where SA_{is} equals one if student i in school s is present at baseline but missing at endline. We test whether γ_1 (differential attrition by treatment status), γ_3 (differential attrition of LBC in treatment schools), or their linear combination differs significantly from zero.

We will also apply the formal tests proposed by Ghanem et al. (2023) to assess attrition bias using baseline data for both respondents and attritors to classify our study as having low or high attrition risk. If we detect differential attrition in either students or teachers, we implement Lee (2009) bounds or appropriate analogues (Ghanem et al., 2024) to assess whether our main results are robust to worst-case assumptions about the missing observations. Given the high rates of cooperation in our partial baseline, we expect attrition to be low.

Missing values and outliers. We check whether patterns of missing data differ systematically across treatment and control groups and by left-behind status using the specifications in Eq. (7)–(8). Observations with missing values on key variables are dropped. For continuous outcomes, we may winsorize at the 1st and 99th percentiles in robustness checks where applicable.

Wild cluster bootstrap. Given the moderate number of clusters (100 schools) in our study, there is ongoing debate about whether cluster-robust standard errors provide adequate inference. As a robustness check, we report wild cluster bootstrap p-values following Cameron et al. (2008), which have been shown to perform well with a limited number of clusters. This serves as a complement to the conventional cluster-robust standard errors reported in our main specifications.

Alternative definition of left-behind status. Our primary analysis defines LBC as students with both parents away from home, representing the most severe form of parental absence. However, children with only one parent absent may also face challenges and be subject to teacher stereotypes. We will test the sensitivity of our results to this definitional choice by re-estimating Eq. (6) using a broader classification: $\text{LBC}_i = 1$ if at least one parent is working away from home, and $\text{LBC}_i = 0$ otherwise.

Nonlinear specifications. Our main specifications use linear models for all outcomes. For bounded or ordinal outcomes—including teacher-assigned grades and Likert scale indices—we pre-specify the following nonlinear robustness checks: probit models for binary outcomes, ordered probit models for ordinal outcomes, and tobit models for censored continuous outcomes. Additionally, following Alesina et al. (2024), we examine the probability of passing (scoring 60 and above) and the probability of getting a good/excellent grade (scoring 80/90 and above) as complementary outcomes to address potential bunching at the two ends of the grade distribution.

5.2 Exploratory Analyses

Compliance and engagement with feedback. We ask teachers to write down their phone number if they want to receive the score for the "categorization test" and all 100 teachers wrote down their phone number during the baseline survey, which is suggestive of 100% take up. As an exploratory test of the awareness channel, we examine whether treatment effects are concentrated among teachers who engaged with the feedback. We use recall accuracy in the endline survey, where treatment teachers are asked to identify the bias category their IAT score fell into. Awareness-driven responses should be larger among teachers who retained the feedback, whereas monitoring or generic-salience channels would produce more uniform effects across engagement.

Difference-in-differences. As exploratory analyses, we estimate the teacher-level difference-in-differences model and the student-level triple-differences model. These specifications allow us to control for individual fixed effects, absorbing all time-invariant teacher and student characteristics respectively. While McKenzie (2012) shows that ANCOVA yields the same coefficient in expectation and provides greater statistical power, these specifications provide an alternative identification strategy.

Teacher-level difference-in-differences. For teacher-level outcomes measured at both baseline and endline, we leverage the panel structure to estimate a difference-in-differences model:

$$Y_{st} = b_0 + b_1 \text{Post}_t + b_2(\text{Feedback}_s \times \text{Post}_t) + \mu_s + v_{st}, \tag{9}$$

where Post_t is an indicator for the endline period ($t = 1$) versus baseline ($t = 0$), and μ_s represents teacher fixed effects. The interaction coefficient b_2 captures the treatment effect on teacher outcomes, while teacher fixed effects absorb all time-invariant teacher characteristics. This specification provides an alternative identification strategy that controls for unobserved heterogeneity across teachers.

Student-level triple-differences. Similarly, we leverage the panel structure of our data to estimate a triple-difference model for outcomes measured at both baseline and endline:

$$\begin{aligned} Y_{ist} = {} & \beta_0 + \beta_1 \text{Post}_t + \beta_2 (\text{Feedback}_s \times \text{Post}_t) \\ & + \beta_3 (\text{LBC}_i \times \text{Post}_t) + \beta_4 (\text{Feedback}_s \times \text{LBC}_i \times \text{Post}_t) + \eta_i + v_{ist}, \end{aligned} \tag{10}$$

where Post_t is an indicator for the endline period ($t = 1$) versus baseline ($t = 0$), and η_i represents student fixed effects. The triple interaction coefficient β_4 captures our parameter of interest—the differential effect of the intervention on LBC relative to non-LBC—while student fixed effects absorb all time-invariant student characteristics. This specification provides an alternative identification strategy that controls for unobserved heterogeneity across students.

Impact heterogeneity. We explore whether treatment effects vary systematically across teacher, school, and student characteristics.

Teacher heterogeneity. We first examine whether effects differ by baseline IAT scores. Two opposing mechanisms are plausible. Teachers with higher baseline bias may exhibit larger treatment effects if they have greater room for bias reduction and experience stronger awareness when confronted with evidence of their implicit associations. Conversely, teachers with lower baseline bias may be more receptive to feedback and more willing to adjust their behavior. Beyond base-

line implicit bias, we test whether treatment effects vary by teacher age, personality, and own left-behind status in childhood.

School heterogeneity. We examine whether effects differ based on the proportion of LBC in the school. In schools with higher LBC concentration, teachers may have more experience with LBC, potentially making the awareness intervention more impactful. Conversely, in schools with lower LBC concentration, LBC constitute a minority and may face more differential treatment, potentially benefiting more from bias reduction. We test for heterogeneous effects by interacting treatment with an indicator for above-median school-level LBC percentage.

Student heterogeneity. We conduct exploratory analyses to examine whether treatment effects vary by student gender, baseline academic performance, and personality. Gender differences may emerge since boys and girls respond differently to teacher biases (Carlana, 2019; Luo and Xie, 2024). The intervention may have differential impacts on high- versus low-performing students. Additionally, students with different personality traits may respond differently to changes in teacher behavior and classroom climate.

Sub-sample analyses by left-behind status. Our main specification uses the interaction term—β_3 in Equation (6)—to capture gap narrowing between LBC and non-LBC. As exploratory analyses, we estimate the following specifications separately on the LBC and non-LBC sub-samples:

$$Y_{is}^1 = \pi_0 + \pi_1 \text{Feedback}_s + X_i'\pi_X + T_s'\pi_T + Z_s'\pi_Z + \omega_{is} \tag{11}$$

For the LBC sub-sample, π_1 captures whether outcomes for LBC improve in absolute terms in treatment schools. Power calculations for this sub-sample are reported in Panel C of Table A2. For the non-LBC sub-sample, π_1 captures how the intervention affects students who are not the direct targets of the bias-reduction message but share the same classroom and teacher.

Channels of the bundled treatment. The intervention comprises of a bundle of personalized IAT feedback and accompanying normative content. Two pre-specified analyses provide partial information about whether the personalized IAT score drives effects or whether the normative message could produce the same response on its own. First, the heterogeneity analysis by baseline IAT score above speaks directly to this distinction: a generic-normative channel does not depend on knowledge of the personalized score and should produce roughly uniform effects across baseline IAT levels, whereas a personalized-information channel should produce effects that vary systematically with baseline IAT scores. Second, the recall analysis above provides a complementary test: if generic normative content were sufficient, effects would not depend on whether teachers correctly recall their specific score; concentration of effects among teachers with accurate recall is

more consistent with the personalized-information channel.

Companion online experiment. We conduct a complementary online experiment with an independent sample of teachers to separately identify the personalized-information channel and the generic-salience channel of the intervention. Extending the supplementary online study in Alesina et al. (2024), we randomize teachers into three arms: (i) personalized IAT feedback combined with normative content on implicit bias toward LBC; (ii) the same normative content without personalized scores; and (iii) no information. The contrast between arms (i) and (ii) isolates the effect of receiving one's own IAT score, holding the normative content constant; the contrast between arms (ii) and (iii) isolates the effect of generic normative content alone. Because the online setting precludes observation of student outcomes, outcomes are limited to teacher-level measures including explicit attitudes toward LBC and the vignette-based picture grading task.

Spillover effects. We explore whether the intervention generates spillover effects beyond the subjects directly taught by the homeroom teacher. If awareness gained by homeroom teachers diffuses to other subject teachers through collegial communication, bias reduction may extend across multiple subjects. We test for these cross-subject spillovers by examining treatment effects on LBC's teacher-graded assignment scores in subjects not taught by the homeroom teacher.

Parental investment mediators. We attempt to collect data on parental remittances and communication frequency from parents of participating students. If successfully collected, we explore whether treatment effects operate through changes in parental investment. Specifically, we examine whether the intervention affects teacher communication with parents and whether this influences parental remittance behavior and engagement with their children's education. This analysis tests whether increased teacher attention to LBC translates into stronger parent-teacher partnerships and enhanced parental involvement, potentially serving as an additional mechanism through which awareness interventions improve student outcomes. However, this analysis is contingent on obtaining adequate parental response rates.

County administrative data and blind grading by out-of-sample teachers. As an exploratory analysis, we will obtain administrative records from non-participating schools in the same county for both the pre-intervention and post-intervention periods from the local Bureau of Education. This allows us to conduct a difference-in-differences comparison between sampled schools (treatment and control combined) and non-sampled schools, providing a benchmark for whether IAT administration itself induces behavioral changes in the control group. A null result—no differential trend between sampled and non-sampled schools—would support the interpretation that

behavioral changes in treatment schools reflect the IAT feedback rather than mere IAT completion.

As an additional exploratory analysis, out-of-sample teachers will blindly grade the same student assignments that teachers evaluate, without knowledge of students' LBC status. Comparing teacher-assigned grades to blind out-of-sample teachers' grades for the same work provides a direct measure of discriminatory grading bias on actual student assignments, holding student performance fixed. A larger gap between teacher-assigned and blind-assigned grades for LBC relative to non-LBC students at baseline, and a narrowing of this gap in treatment schools at endline, would provide direct evidence of reduced discriminatory grading. This analysis is subject to the feasibility of collecting and anonymizing student assignments at scale.

Follow up into next academic year. We aim to conduct follow-up data collection at the start of the next academic year (November 2026), tracking the same students in Grade 6 and their homeroom teachers. In Chinese elementary schools, homeroom teachers typically remain with the same cohort across grades, meaning that the treated teacher is likely to remain the primary point of contact for the same students in the large majority of cases. We expect teacher transitions to affect only a small share of schools (approximately 10%), primarily due to school consolidation, and we will exclude these schools from the follow-up analysis. The follow-up will collect the same set of outcomes as the endline survey for both teachers and students, allowing us to test whether treatment effects persist, strengthen, or fade over time—both for direct outcomes such as teacher attitudes and grading behavior, and for downstream outcomes such as peer relationships, school sense of belonging, behavioral problems, self-confidence, and blindly graded standardized test scores. We acknowledge that the follow-up sample will be somewhat smaller due to necessary exclusions. To preserve the treatment-control contrast at follow-up, we also plan to delay the IAT feedback for control teachers until follow-up data collection is completed (approximately December 2026). This exploratory analysis is subject to obtaining the necessary IRB extensions and approval from the local education bureau. If the follow-up cannot be conducted, null results on downstream outcomes at endline will be interpreted as potentially reflecting insufficient exposure rather than evidence against the theory of change, given the short 6–8 week intervention window.

6 Preliminary Validation from Partial Baseline

At the time of PAP submission, we collected baseline data from 21 of the 100 target schools. Baseline data collection resumes in March after the school break. This section presents preliminary descriptive evidence on teacher bias toward LBC using this partial sample. Specifically, we partially test Hypothesis 1 to provide suggestive evidence of existing biases and thus the need for our bias-awareness intervention. Once the study is completed and the full sample is available, this

section will be replaced with results from all 100 schools.

6.1 Teacher Implicit Bias at Baseline

Figure 2 presents the distribution of IAT D-scores from the 21-school baseline sample. The distribution shows considerable variation in implicit bias, with scores ranging from approximately -0.7 to 1.3. The majority of teachers exhibit positive D-scores, indicating negative implicit bias against LBC. A substantial proportion fall above the 0.35 threshold, indicating moderate to strong bias.

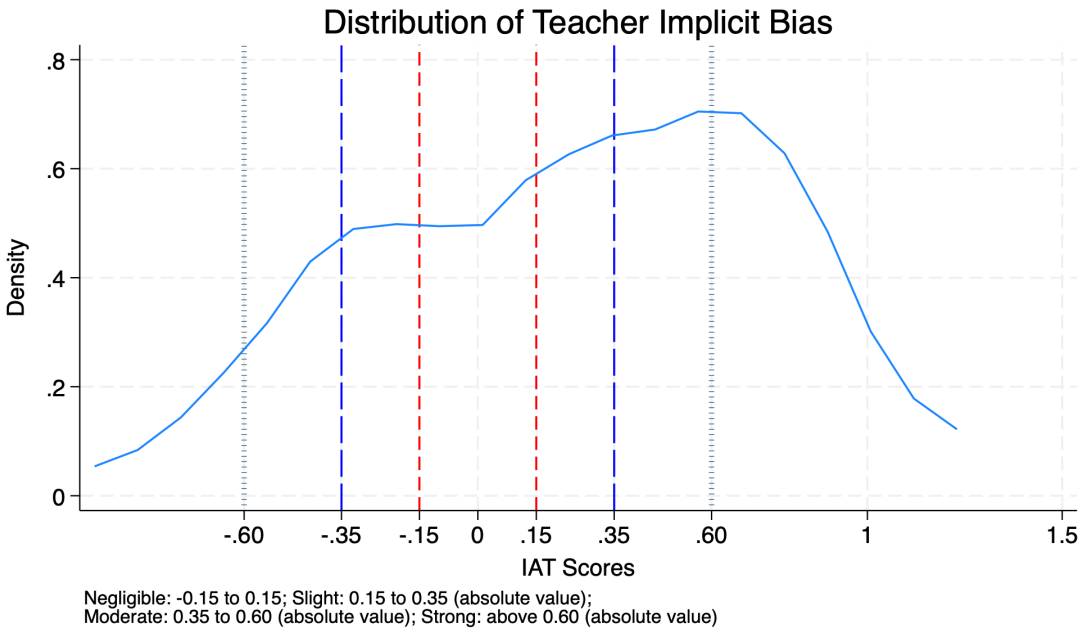


Figure 2: Distribution of the LBC IAT Score across Teachers

6.2 Implicit Bias and Grading Behavior

To examine whether implicit bias affects grading decisions, we introduce a simple picture grading task in the teacher survey. Teachers evaluate AI-generated children's drawings without being told the study's focus on LBC. We present two sets of drawings that implicitly signal left-behind versus non-left-behind status through depicted life scenarios—for example, living with grandparents (LBC) versus living with parents (non-LBC). Teachers assign grades to each drawing on a standardized scale (see Appendix B.4).

Figure 3 shows the relationship between teachers' implicit bias (IAT scores) and the grading gap, calculated as the difference between grades assigned to non-left-behind drawings minus grades assigned to left-behind drawings. Each point represents one teacher. The positive slope

indicates that teachers with stronger implicit bias assign systematically lower grades to drawings depicting left-behind life scenarios relative to non-left-behind scenarios.

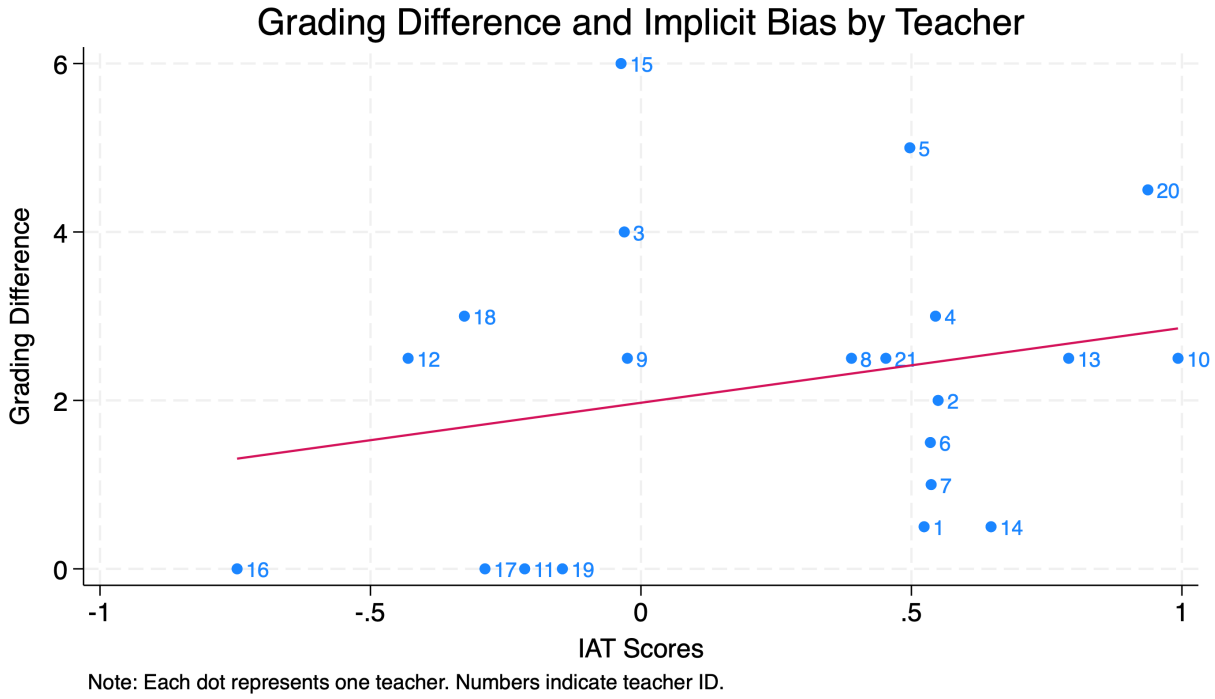


Figure 3: Grading Difference versus Implicit Bias

6.3 Implicit Bias and Explicit Attitudes

Figure 4 examines the correlation between implicit (IAT) and explicit bias measures. Higher values of explicit attitude indicate more negative attitudes toward classrooms with higher proportions of LBC. The positive correlation suggests that teachers with stronger implicit bias against LBC also express more negative explicit attitudes. This pattern gives us confidence in the validity of our measures.

6.4 Implicit Bias and Student Outcomes

We examine whether teachers' implicit bias is associated with disparities between LBC and non-LBC across multiple dimensions at baseline. For all outcomes, we estimate Eq. (4) with teacher fixed effects, where the coefficient on the interaction term $(\text{IAT}_s \times \text{LBC}_i)$ captures whether outcome gaps vary systematically with teacher bias.

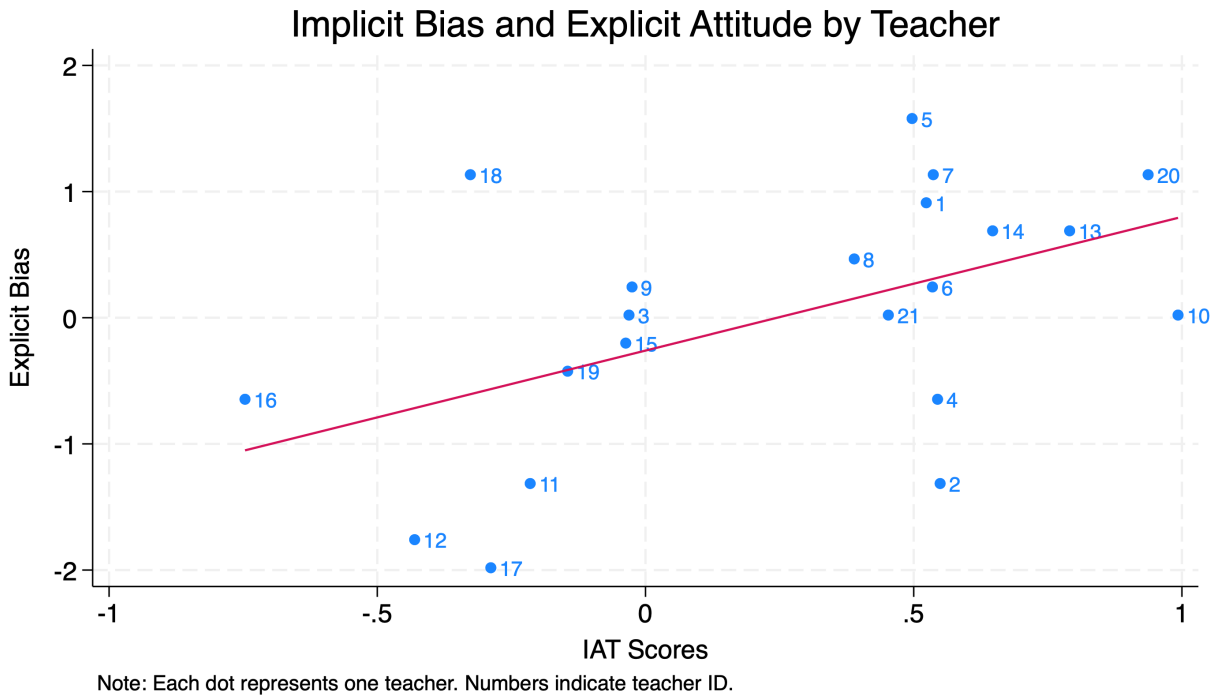


Figure 4: Explicit Bias versus Implicit Bias

Academic Outcomes. We examine whether teacher implicit bias correlates with student achievement using self-reported grades in math and Chinese. At the time of PAP submission, first-term examinations had just been completed, and results were not yet available. We therefore rely on students' self-reported grades from the previous academic year (Grade 4, second term). Students report scores in categorical ranges (below 60, 60-70, 70-80, 80-90, 90-100), which we convert to interval midpoints and standardize. We acknowledge that self-reported grades may be subject to measurement error and social desirability bias. Once baseline data collection is complete, we will test Hypothesis 1(b) using teacher-assigned grades and blindly graded standardized test scores from administrative records.

Table 2 presents preliminary correlational evidence. The interaction coefficient captures the association between teacher implicit bias and the LBC achievement gap. For math scores (Column 1), the coefficient is negative, suggesting that in classrooms where teachers hold stronger implicit bias, LBC report lower grades relative to non-LBC peers. For Chinese scores (Column 2), the interaction coefficient is also negative but smaller in magnitude.

These preliminary patterns are consistent with teacher implicit bias being associated with achievement disparities, though several important caveats apply. First, the reliance on self-reported grades introduces measurement error that may attenuate true effects. Second, the small sample size (21 schools, approximately 700 students) provides limited statistical power so we refrain from

Table 2: Teacher Implicit Bias and Student Academic Outcome

	(1) Math Score	(2) Chinese Score
LBC	0.052	-0.043
	(0.110)	(0.114)
IAT × *LBC*	-0.429**	-0.059
	(0.193)	(0.214)
Observations	700	703
R-squared	0.168	0.229
Teacher FE	Yes	Yes
Student controls	Yes	Yes
Dep. Mean	0.000	0.000

Notes: The table reports associations between teachers' implicit bias and student academic outcomes at baseline. Dependent variables are standardized math and Chinese scores based on students' self-reported grades from the previous semester's blindly graded examinations. All regressions include teacher fixed effects and student-level controls: gender, only-child status, and categorical indicators for mother's and father's education (completed compulsory education, high school or above, unknown). Standard errors clustered at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

making inferences based on statistical significance. Third, these are correlational patterns from baseline data and do not establish causality. The randomized intervention will provide causal evidence on whether reducing teacher bias narrows achievement gaps using objective measures.

Peer Relationships. Table 3 examines the association between teacher implicit bias and student peer relationships. The interaction coefficients suggest that in classrooms with more biased teachers, LBC experience greater social exclusion: they report fewer friends, nominate fewer peers as friends, and are more likely to make no friendship nominations. These patterns indicate that teacher implicit bias is associated with the social marginalization of LBC.

We also examine other non-academic outcomes including behavioral problems, self-confidence and expectations, and sense of belonging at school (not reported; available upon request). While we find suggestive patterns indicating that LBC in high-bias classrooms experience worse outcomes across these dimensions, the estimates are imprecise given the limited baseline sample. Nevertheless, these promising preliminary findings motivate our experimental intervention to test whether reducing teacher bias can improve educational experiences for LBC across multiple domains.

7 Expected Findings

Teacher bias represents a critical yet underexplored mechanism perpetuating educational inequality for LBC. Through this RCT, we provide the first experimental evidence on whether low-cost awareness interventions can reduce teacher bias toward this disadvantaged population and produce

Table 3: Teacher Implicit Bias and Student Peer Relationships

	(1) Number of Friends	(2) Out-degree Continuous	(3) In-degree Continuous	(4) Out-degree Binary	(5) In-degree Binary
LBC	-0.017	0.103	-0.428	-0.026	-0.003
	(0.683)	(0.228)	(0.358)	(0.021)	(0.045)
IAT $\times$ *LBC*	-2.649**	-0.756*	0.336	0.090**	0.054
	(1.015)	(0.364)	(0.719)	(0.033)	(0.085)
Observations	721	721	721	721	721
R-squared	0.120	0.112	0.043	0.085	0.0516
Teacher FE	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes
Dep. Mean	6.541	3.959	3.742	0.053	0.084

Notes: The table reports associations between teachers' implicit bias and student peer relationships at baseline. Column 1 reports the self-reported number of friends in the classroom. Columns 2-3 report continuous measures of out-degree (number of friendship nominations made) and in-degree (number of nominations received). Columns 4-5 report binary indicators for social isolation: in-degree binary equals one if the student receives no nominations; out-degree binary equals one if the student makes no nominations. All regressions include teacher fixed effects and student-level controls: gender, only-child status, and categorical indicators for mother's and father's education (completed compulsory education, high school or above, unknown). Standard errors clustered at the school level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

comprehensive educational improvements beyond subjective evaluations. We expect the intervention to operate through a clear causal chain. First, awareness feedback reduces teachers' explicit bias toward LBC, confirming that the intervention successfully increases awareness. Second, the intervention reduces discriminatory grading and improves daily classroom interactions that build confidence and belonging. These changes in teacher behavior generate downstream improvements in LBC's behaviors, well-being, and peer relationships—outcomes that are especially critical for a population that experiences not only achievement gaps but also social isolation and psychological distress stemming from parental absence. Together, fairer grading, improved classroom interactions, and enhanced student well-being demonstrate that awareness interventions can produce comprehensive educational improvements beyond teacher-controlled evaluations.

If the intervention successfully reduces bias and improves educational experiences for LBC, this would demonstrate that scalable, low-cost awareness interventions can address structural disadvantages facing vulnerable student populations. Unlike prior studies examining bias based on gender or immigration status, our findings would show that awareness interventions can work even for silent populations whose disadvantage lacks observable markers: LBC share the same appearance and spoken language as their peers but experience parental absence. This has important implications for education policy: many vulnerable student groups, including those experiencing poverty, family instability, or other forms of hidden disadvantage, may benefit from similar inter-

ventions. Given the global prevalence of labor migration and LBC across Southeast Asia, Latin America, Eastern Europe, and sub-Saharan Africa, these findings could inform evidence-based policies to reduce educational inequality in diverse contexts where parental migration creates challenges for children's development.

References

Alan, Sule, Enes Duysak, Elif Kubilay, and Ipek Mumcu, "Social exclusion and ethnic segregation in schools: The role of teachers' ethnic prejudice," *Review of Economics and Statistics*, September 2023, *105* (5), 1039–1054.

_ , Michela Carlana, and Marinella Leone, "Inclusive teaching: Spotting social isolation in the classroom," National Bureau of Economic Research 2024.

_ , Seda Ertac, and Ipek Mumcu, "Gender stereotypes in the classroom and effects on achievement," *The Review of Economics and Statistics*, 2018, *100* (5), 876–890.

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti, "Revealing stereotypes: Evidence from immigrants in schools," *American Economic Review*, July 2024, *114* (7), 1916–1948.

Alwin, Duane F and Jon A Krosnick, "The reliability of survey attitude measurement: The influence of question and respondent attributes," *Sociological methods & research*, 1991, *20* (1), 139–181.

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

Banerjee, Ritwik, Satarupa Mitra, Soham Sahoo, and Ashmita Gupta, "Caste identity and teachers' biased expectations: Evidence from Bihar, India," *Journal of Development Economics*, 2025, p. 103650.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 2006, *93* (3), 491–507.

Blanton, Hart, James Jaccard, Jonathan Klick, Barbara Mellers, Gregory Mitchell, and Philip E Tetlock, "Strong claims and weak evidence: reassessing the predictive validity of the IAT.," *Journal of applied Psychology*, 2009, *94* (3), 567.

Boring, Anne and Arnaud Philippe, "Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching," *Journal of Public Economics*, January 2021, *193*, 104323.

Botelho, Fernando, Ricardo A. Madeira, and Marcos A. Rangel, "Racial discrimination in grading: Evidence from Brazil," *American Economic Journal: Applied Economics*, 2015, *7* (4), 37–52.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American economic journal: applied economics*, 2009, *1* (4), 200–232.

Burgess, Simon and Ellen Greaves, "Test scores, subjective assessment, and stereotyping of ethnic minorities," *Journal of Labor Economics*, 2013, *31* (3), 535–576.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott, "Misperceived social norms: Women working outside the home in Saudi Arabia," *American Economic Review*, October 2020, *110* (10), 2997–3029.

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

Carlana, Michela, "Implicit stereotypes: Evidence from teachers' gender bias," *The Quarterly Journal of Economics*, August 2019, *134* (3), 1163–1224.

_ , Eliana La Ferrara, and Paolo Pinotti, "Implicit stereotypes in teachers' track recommendations," *AEA Papers and Proceedings*, May 2022, *112*, 409–414.

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

Dee, Thomas S, "A teacher like me: Does race, ethnicity, or gender matter?," *American Economic Review*, 2005, *95* (2), 158–165.

Fellmeth, Gracia, Kelly Rose-Clarke, Chenyue Zhao, Laura K Busert, Yunting Zheng, Alessandro Massazza, Hacer Sonmez, Ben Eder, Alice Blewitt, Wachiraya Lertgrai, Miriam Orcutt, Katharina Ricci, Olaa Mohamed-Ahmed, Rachel Burns, Duleeka Knipe,

Sally Hargreaves, Therese Hesketh, Charles Opondo, and Delan Devakumar, "Health impacts of parental migration on left-behind children and adolescents: A systematic review and meta-analysis," *The Lancet*, December 2018, *392* (10164), 2567–2582.

Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Becerra, "Testing attrition bias in field experiments," *Journal of Human Resources*, 2023, *58* (3), 1010–1047.

_ , _ , Désiré Kédagni, and Karen Ortiz-Becerra, "Correcting attrition bias using changes-in-changes," *Journal of Econometrics*, 2024, *105737*.

Glover, Dylan, Amanda Pallais, and William Pariente, "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores," *The Quarterly Journal of Economics*, 2017, *132* (3), 1219–1260.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L. K. Schwartz, "Measuring individual differences in implicit cognition: The Implicit Association Test.," *Journal of Personality and Social Psychology*, June 1998, *74* (6), 1464–1480.

Greenwald, Anthony G, Miguel Brendl, Huajian Cai, Dario Cvencek, John F Dovidio, Malte Friese, Adam Hahn, Eric Hehman, Wilhelm Hofmann, Sean Hughes et al., "Best research practices for using the Implicit Association Test," *Behavior Research Methods*, 2022, *54* (3), 1161–1180.

_ , T Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R Banaji, "Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity.," *Journal of Personality and Social Psychology*, 2009, *97* (1), 17.

Hanna, Rema N. and Leigh L. Linden, "Discrimination in grading," *American Economic Journal: Economic Policy*, 2012, *4* (4), 146–168.

Heckman, James J., "Skill formation and the economics of investing in disadvantaged children," *Science*, 2006, *312* (5782), 1900–1902.

Huang, Zibin and Junsen Zhang, "School restrictions, migration, and peer effects: A spatial equilibrium analysis of children's human capital in China," 2025.

Hunan Provincial Bureau of Statistics, "Statistical bulletin on economic and social development of Hunan province in 2023," http://tjj.hunan.gov.cn/hntj/tjfx/tjgb/jjfzgb/202403/t20240322_33260459.html 2024.

Karing, Constance, Tobias Rausch, and Cordula Artelt, "Teacher judgement accuracy—measurements, causes and effects," in "Educational Processes, Decisions, and the Development of Competencies from Early Preschool Age to Adolescence: Findings from the BiKS Cohort Panel Studies," Springer, 2024, pp. 263–280.

Kerwin, Jason, Nada Rostom, and Olivier Sterck, "Striking the right balance: Why standard balance tests over-reject the null, and how to fix it," 2024.

Lavy, Victor and Edith Sand, "On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases," *Journal of Public Economics*, November 2018, *167*, 263–279.

_ **and Rigissa Megalokonomou**, "The short-and the long-run impact of gender-biased teachers," *American Economic Journal: Applied Economics*, 2024, *16* (2), 176–218.

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

Li, Xinyu, Wei Jin, Lu Han, Xingyu Chen, and Lihong Li, "Comparison and application of depression screening tools for adolescents: scale selection and clinical practice," *Child and Adolescent Psychiatry and Mental Health*, May 2025, *19* (1), 53.

Luo, Qinyue and Huihua Xie, "Test scores, noncognitive outcomes, and the stereotyping of non-local students," RF Berlin-CReAM Discussion Paper Series 2024.

McKenzie, David, "Beyond baseline and follow-up: The case for more T in experiments," *Journal of Development Economics*, 2012, *99* (2), 210–221.

Mengel, Friederike, Jan Sauermann, and Ulf Zolitz, "Gender bias in teaching evaluations," *Journal of the European Economic Association*, 2019, *17* (2), 535–566.

National Bureau of Statistics of China, "Residents' income and consumption expenditure in 2023," https://www.stats.gov.cn/sj/zxfb/202401/t20240116_1946622.html 2024.

NBS, UNICEF, and UNFPA, "What the 2020 census can tell us about children in China: Facts and figures," Technical Report, National Bureau of Statistics of China, UNICEF China, UNFPA China 2023.

Oswald, Frederick L., Gregory Mitchell, Hart Blanton, James Jaccard, and Philip E. Tetlock, "Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies.," *Journal of Personality and Social Psychology*, August 2013, *105* (2), 171–192.

Papageorge, Nicholas W., Seth Gershenson, and Kyung Min Kang, "Teacher expectations matter," *The Review of Economics and Statistics*, May 2020, *102* (2), 234–251.

Rakshit, Sonali and Soham Sahoo, "Biased teachers and gender gap in learning outcomes: Evidence from India," *Journal of Development Economics*, 2023, *161*, 103041.

Ramachandran, Rajesh, Devesh Rustagi, and Emilia Soldani, "Discrimination by teachers: role of attitudes, beliefs, and empathy," 2025.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales, "How stereotypes impair women's careers in science," *Proceedings of the National Academy of Sciences of the United States of America*, 2014, *111* (12), 4403–4408.

Rooth, Dan-Olof, "Automatic associations and discrimination in hiring: Real world evidence," *Labour Economics*, June 2010, *17* (3), 523–534.

Torsheim, Torbjoen, Bente Wold, and Oddrun Samdal, "The teacher and classmate support scale: factor structure, test-retest reliability and validity in samples of 13-and 15-year-old adolescents," *School Psychology International*, 2000, *21* (2), 195–212.

UNICEF, "Children Left-Behind," Technical Report, United Nations Children's Fund (UNICEF), New York 2025.

Yue, Ai, Yu Bai, Yaojiang Shi, Renfu Luo, Scott Rozelle, Alexis Medina, and Sean Sylvia, "Parental migration and early childhood development in rural China," *Demography*, 2020, *57* (2), 403–422.

Zhang, Hongliang, Jere R. Behrman, C. Simon Fan, Xiangdong Wei, and Junsen Zhang, "Does parental absence reduce cognitive achievements? Evidence from rural China," *Journal of Development Economics*, 2014, *111*, 181–195.

Zhou, Chi, Qiaohong Lv, Nancy Yang, and Feng Wang, "Left-behind children, parent-child communication and psychological resilience: A structural equation modeling analysis," *International Journal of Environmental Research and Public Health*, 2021, *18* (10), 5123.

A Appendix A: A Stylized Model

This section presents a simple framework that formalizes the Theory of Change and generates the testable predictions outlined in Section 2.2. Following the structure in Figure 1, the model captures how IAT awareness feedback affects student outcomes through: (1) increased teacher awareness (the mechanism), (2) direct effects on teacher behaviors (grading and interactions), and (3) downstream effects on student wellbeing and learning.

A.1 Setup

Consider a classroom with teacher s and student i. Each student has innate ability θ_i, which is unobserved by the teacher. Students differ in their family structure, with $\text{LBC}_i = 1$ indicating a left-behind child and $\text{LBC}_i = 0$ indicating a non-left-behind child.

A.1.1 Student Achievement

Student i's performance on blind-graded standardized tests is given by the production function:

$$T_{is} = f(\theta_i, c_{is}, e_{is}) \tag{A1}$$

where:

- θ_i is student i's innate ability
- c_{is} is student i's confidence and motivation
- e_{is} is the classroom learning environment

We assume f is increasing in all its arguments: $f_\theta > 0$, $f_c > 0$, $f_e > 0$.

A.1.2 Confidence Formation

Student confidence is shaped by the feedback students receive through teacher-assigned grades:

$$c_{is} = g(G_{is}, x_i) \tag{A2}$$

where G_{is} represents teacher-assigned grades and x_i captures student characteristics. We assume $g_G > 0$: higher teacher-assigned grades increase student confidence and motivation. This captures the well-documented feedback effect of teacher evaluations on student self-perception (Papageorge et al., 2020).

A.1.3 Teacher-Assigned Grades

Teachers assign grades based on their perception of student ability, but these evaluations may be influenced by implicit bias:

$$G_{is} = h(\theta_i, \text{LBC}_i, b_s) \tag{A3}$$

where b_s is teacher s's bias parameter. We assume:

- $h_\theta > 0$: higher ability leads to higher grades
- $h(\theta, 1, b_s) - h(\theta, 0, b_s) < 0$ when $b_s > 0$: biased teachers assign lower grades to LBC conditional on ability

For concreteness, one can think of $h(\theta_i, \text{LBC}_i, b_s) = \tilde{h}(\theta_i) - b_s \cdot \text{LBC}_i$, where $b_s > 0$ represents a direct grade penalty for LBC.

A.1.4 Classroom Learning Environment

The classroom learning environment depends on teacher quality and effort, which may vary based on teacher bias:

$$e_{is} = k(\text{LBC}_i, b_s, x_s) \tag{A4}$$

where x_s represents other teacher and school characteristics. We assume:

- $k_b < 0$ for $\text{LBC}_i = 1$: teachers with higher bias provide lower quality instruction or fewer resources to LBC

A.2 The IAT Awareness Intervention

The intervention provides teachers with personalized feedback on their IAT scores, increasing awareness of implicit bias. Let $\text{Treatment}_s \in \{0, 1\}$ denote random assignment at the school level, where $\text{Treatment}_s = 1$ indicates the treatment group (immediate IAT feedback) and $\text{Treatment}_s = 0$ indicates the control group (delayed feedback).

The intervention operates by reducing teachers' bias parameter. Once aware of their implicit bias, teachers in the treatment group reduce discriminatory behavior in grading and classroom interactions:

$$b_s(\text{Treatment}_s) = \begin{cases} b_s^0 \cdot (1 - \delta) & \text{if } \text{Treatment}_s = 1 \\ b_s^0 & \text{if } \text{Treatment}_s = 0 \end{cases} \tag{A5}$$

where b_s^0 is the teacher's baseline bias level and $\delta \in (0, 1)$ measures the effectiveness of the awareness intervention in reducing bias.

This bias reduction produces two direct effects. First, teacher-assigned grades become less discriminatory. For students of equal ability θ, the grading gap is:

$$h(\theta, 0, b_s) - h(\theta, 1, b_s)$$

Since treatment reduces b_s, the grading gap narrows in the treatment group.

Second, the classroom environment improves for LBC:

$$e_{is}(\text{Treatment}_s) = k(\text{LBC}_i, b_s(\text{Treatment}_s), x_s) \tag{A6}$$

Since $k_b < 0$ for $\text{LBC}_i = 1$ and $b_s(\text{Treatment}_s = 1) < b_s(\text{Treatment}_s = 0)$, we have:

$$e_{is}(\text{Treatment}_s = 1) > e_{is}(\text{Treatment}_s = 0) \quad \text{for LBC} \tag{A7}$$

A.3 Testable Predictions

We derive comparative statics that generate our pre-specified hypotheses. The predictions follow the structure of our Theory of Change: baseline bias (Hypothesis 1), mechanism (Hypothesis 2), direct outcomes (Hypotheses 3–4), and downstream outcomes (Hypotheses 5–6).

A.3.1 Hypothesis 1: Baseline Bias

H1(a): Teachers exhibit implicit bias. At baseline, we expect:

$$\mathbb{E}[b_s^0] > 0 \tag{A8}$$

Empirical Test: Positive average IAT D-scores at baseline, indicating teachers associate LBC with negative attributes.

H1(b): Implicit bias is associated with worse outcomes for LBC. In classrooms where teachers hold stronger baseline bias, the gap between LBC and non-LBC should be larger. For academic performance:

$$\frac{\partial}{\partial b_s^0}\left[\mathbb{E}[G_{is}|\text{LBC}_i = 1] - \mathbb{E}[G_{is}|\text{LBC}_i = 0]\right] < 0 \tag{A9}$$

Similarly for non-academic outcomes (wellbeing, peer relationships, engagement), higher teacher bias should be associated with larger gaps.

Empirical Test: In baseline regressions, the interaction between IAT scores and LBC status predicts outcome gaps. Higher baseline IAT scores are associated with larger LBC-non-LBC gaps in academic performance and non-academic outcomes.

A.3.2 Hypothesis 2: Awareness and Explicit Attitudes (The Mechanism)

The intervention operates by increasing teachers' awareness of their implicit bias, which leads to the reduction in the bias parameter b_s modeled above. We test this mechanism empirically by measuring changes in teachers' explicit attitudes toward LBC.

Empirical Test: Treatment reduces teachers' negative explicit attitudes toward classrooms with higher proportions of LBC, confirming that the awareness channel is activated.

A.3.3 Hypothesis 3: Direct Effect on Grading

The intervention's effect on grading operates through a difference-in-differences mechanism. Treatment reduces grading bias, which affects LBC and non-LBC differentially.

For non-LBC:

$$\text{TE}_{\text{grades}}^{\text{non-LBC}} = \mathbb{E}[G_{is}(\text{Treatment}_s = 1)|\text{LBC}_i = 0] - \mathbb{E}[G_{is}(\text{Treatment}_s = 0)|\text{LBC}_i = 0] \tag{A10}$$

For LBC, treatment increases grades by reducing discriminatory bias:

$$\begin{aligned}\text{TE}_{\text{grades}}^{\text{LBC}} &= \mathbb{E}[G_{is}(\text{Treatment}_s = 1)|\text{LBC}_i = 1] - \mathbb{E}[G_{is}(\text{Treatment}_s = 0)|\text{LBC}_i = 1] \\ &= \mathbb{E}[h(\theta_i, 1, b_s^0(1-\delta)) - h(\theta_i, 1, b_s^0)|\text{LBC}_i = 1] > 0\end{aligned} \tag{A11}$$

Since treatment specifically reduces bias against LBC (reducing b_s), we predict treatment effects are larger for LBC. The key prediction is the differential treatment effect:

$$\text{DTE}_{\text{grades}} = \text{TE}_{\text{grades}}^{\text{LBC}} - \text{TE}_{\text{grades}}^{\text{non-LBC}} > 0 \tag{A12}$$

This measures how much more treatment affects LBC compared to non-LBC. Treatment should narrow the grading gap:

$$\underbrace{[\mathbb{E}[G_{is}^{\text{non-LBC}}] - \mathbb{E}[G_{is}^{\text{LBC}}]]}_{\text{Gap in control}} > \underbrace{[\mathbb{E}[G_{is}^{\text{non-LBC}}] - \mathbb{E}[G_{is}^{\text{LBC}}]]}_{\text{Gap in treatment}} \tag{A13}$$

Empirical Test: To test whether the intervention reduces discriminatory grading, we compare teacher-assigned grades for LBC and non-LBC students across treatment and control schools. If the intervention reduces bias in grading, we expect the gap to be smaller in treatment schools, indicated by a positive differential treatment effect.

A.3.4 Hypothesis 4: Direct Effect on Teacher-Student Interactions

The treatment's effect on teaching engagement and teacher-student interactions follows a difference-in-differences structure. While not separately modeled above for parsimony, teacher-student interactions are a key component of the classroom learning environment (e_{is}) described in Eq. (A4). We empirically measure perceived interaction quality through student reports. Let I_{is} denote perceived interaction quality for student i in school s.

For non-LBC:

$$\text{TE}^{\text{non-LBC}}_{\text{interactions}} = \mathbb{E}[I_{is}|\text{LBC}_i = 0, \text{Treatment}_s = 1] - \mathbb{E}[I_{is}|\text{LBC}_i = 0, \text{Treatment}_s = 0] \tag{A14}$$

For LBC, treatment improves perceived interactions through enhanced classroom environment:

$$\text{TE}^{\text{LBC}}_{\text{interactions}} = \mathbb{E}[I_{is}|\text{LBC}_i = 1, \text{Treatment}_s = 1] - \mathbb{E}[I_{is}|\text{LBC}_i = 1, \text{Treatment}_s = 0] > 0 \tag{A15}$$

Since treatment reduces bias specifically against LBC, we predict larger treatment effects for LBC. The differential treatment effect:

$$\text{DTE}_{\text{interactions}} = \text{TE}^{\text{LBC}}_{\text{interactions}} - \text{TE}^{\text{non-LBC}}_{\text{interactions}} > 0 \tag{A16}$$

This narrows the gap in perceived teacher-student interactions between LBC and non-LBC.

Empirical Test: To test whether the intervention improves classroom interactions, we compare how LBC and non-LBC students perceive their interactions with teachers across treatment and control schools. If the intervention improves teaching engagement toward LBC, we expect the gap in perceived teacher-student interactions to be smaller in treatment schools, indicated by a positive differential treatment effect.

A.3.5 Hypothesis 5: Downstream Effect on Student Well-being and Behaviors

Through improved teacher-student interactions, the intervention enhances students' non-academic outcomes. While the model focuses on how confidence (c_{is}) and learning environment (e_{is}) affect test scores, these factors also influence broader dimensions of student well-being and social integration that we measure empirically. Let W^j_{is} denote outcome dimension $j \in \{\text{well-being, peer relationships}\}$ for student i.

For non-LBC:

$$\text{TE}^{\text{non-LBC}}_{j} = \mathbb{E}[W^j_{is}|\text{LBC}_i = 0, \text{Treatment}_s = 1] - \mathbb{E}[W^j_{is}|\text{LBC}_i = 0, \text{Treatment}_s = 0] \tag{A17}$$

For LBC, treatment improves outcomes through the enhanced classroom environment:

$$\text{TE}_j^{\text{LBC}} = \mathbb{E}[W_{is}^j | \text{LBC}_i = 1, \text{Treatment}_s = 1] - \mathbb{E}[W_{is}^j | \text{LBC}_i = 1, \text{Treatment}_s = 0] > 0 \tag{A18}$$

Since the improvements in teacher behavior (Hypotheses 3 and 4) primarily benefit LBC, downstream effects on wellbeing should also be larger for LBC. The differential treatment effect:

$$\text{DTE}_j = \text{TE}_j^{\text{LBC}} - \text{TE}_j^{\text{non-LBC}} > 0 \tag{A19}$$

Empirical Test: To test whether the intervention improves student well-being and peer relationships, we compare these outcomes for LBC and non-LBC students across treatment and control schools. If improved teacher interactions translate into better social and emotional outcomes for LBC, we expect gaps to be smaller in treatment schools, indicated by positive differential treatment effects.

A.3.6 Hypothesis 6: Downstream Effect on Learning Outcomes

Through improved grading behavior, teacher-student interactions and enhanced student well-being, the intervention ultimately affects blindly graded standardized test scores.

For non-LBC:

$$\text{TE}_{\text{tests}}^{\text{non-LBC}} = \mathbb{E}[T_{is}(\text{Treatment}_s = 1) | \text{LBC}_i = 0] - \mathbb{E}[T_{is}(\text{Treatment}_s = 0) | \text{LBC}_i = 0] \tag{A20}$$

For LBC, treatment improves test scores through two downstream channels:

$$\begin{aligned} \text{TE}_{\text{tests}}^{\text{LBC}} &= \mathbb{E}[T_{is}(\text{Treatment}_s = 1) | \text{LBC}_i = 1] - \mathbb{E}[T_{is}(\text{Treatment}_s = 0) | \text{LBC}_i = 1] \\ &= \mathbb{E}\left[f\left(\theta_i, c_{is}(G_{is}^{T=1}), e_{is}^{T=1}\right) - f\left(\theta_i, c_{is}(G_{is}^{T=0}), e_{is}^{T=0}\right) \middle| \text{LBC}_i = 1 \right] > 0. \end{aligned} \tag{A21}$$

Since both channels (grading and classroom environment) primarily affect LBC, we predict larger treatment effects on test scores for LBC. The differential treatment effect:

$$\text{DTE}_{\text{tests}} = \text{TE}_{\text{tests}}^{\text{LBC}} - \text{TE}_{\text{tests}}^{\text{non-LBC}} > 0 \tag{A22}$$

Empirical Test: To test whether the intervention improves learning outcomes, we compare blindly graded standardized test scores for LBC and non-LBC students across treatment and control schools. If the intervention generates downstream improvements in learning through the channels described above (improved grading, teacher interactions, and student well-being), we expect the achievement gap to be smaller in treatment schools, indicated by a positive differential treatment effect.

B Appendix B: Additional Materials

B.1 Description of the IAT

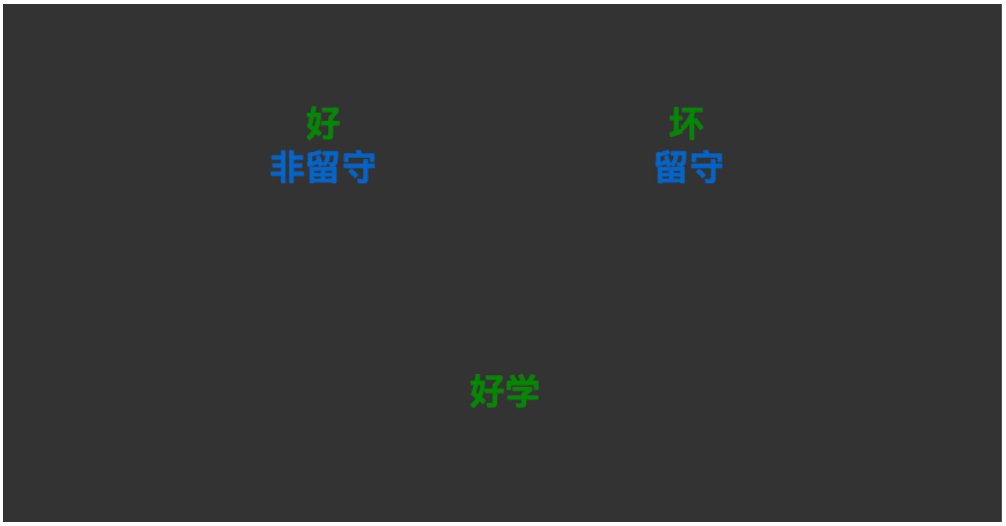
The IAT for this study measures implicit associations between LBC and evaluative attributes using two sets of stimuli. The first set contrasts LBC with non-LBC using descriptive phrases. Left-behind stimuli include phrases such as "Parents away" and "Raised by grandparents." Non-left-behind stimuli include phrases such as "Parental care" and "Parent-child cohabitation." The second set consists of positive adjectives (e.g., intelligent, capable) and negative adjectives (e.g., disrespectful, lazy). Teachers complete the IAT on a computer. During the test, category labels are displayed at the top corners of the screen, while target and attribute words appear randomly in the center. Teachers are instructed to classify each word into the appropriate category as rapidly as possible by pressing designated keys. In one type of block, teachers categorize left-behind-related phrases and negative adjectives to the same key. In another type of block, left-behind-related phrases pair with positive adjectives instead. Figure A1 shows the example of the screenshot in the compatible task (the test was administered in Chinese; illustration translated to English).

The IAT consists of seven blocks following standard protocols. Blocks 3-4 and 6-7 are the critical test blocks: "compatible" blocks pair left-behind with negative attributes (consistent with stereotypes), while "incompatible" blocks pair left-behind with positive attributes (inconsistent with stereotypes). Block order is randomized at the individual level: half of teachers complete compatible blocks first (1, 2, 3, 4, 5, 6, 7; as shown in Table A1), while the other half complete an alternative sequence starting with incompatible pairings (1, 5, 6, 7, 2, 3, 4). D-scores are calculated using response times from blocks 3, 4, 6, and 7, following the improved scoring algorithm (Greenwald et al., 2022) as detailed below. Higher D-scores indicate stronger associations between LBC and negative attributes.

IAT Stimuli for Left-Behind Children Study

1. **Left-behind**: Left-behind, Parents away, Parents absent, Relative care, Raised by grandparents, Live apart from parents
2. **Non-left-behind**: Non-left-behind, Parents at home, Parental care, Parental supervision, Parent-child cohabitation, Live with parents
3. **Good**: Intelligent, Capable, Studious, Attentive, Good-willed, Respectful
4. **Bad**: Disrespectful, Incapable, Boisterous, Lazy, Distracted, Sloppy

(a) English version



(b) Chinese version

Figure A1: IAT Block Structure (Compatible-First Sequence: 1, 2, 3, 4, 5, 6, 7)

Blocks	Left Categories	Right Categories
1	Non-left-behind	Left-behind
2	Good	Bad
3 (compatible)	Non-left-behind + Good	Left-behind + Bad
4 (compatible)	Non-left-behind + Good	Left-behind + Bad
5	Bad	Good
6 (incompatible)	Non-left-behind + Bad	Left-behind + Good
7 (incompatible)	Non-left-behind + Bad	Left-behind + Good

Table A1: IAT Block Structure (compatible blocks first sequence)

IAT Scoring Algorithm

Data Exclusion Criteria:

1. All practice trials from Blocks 1, 2, and 5 were discarded.
2. Trials with latencies greater than 10,000 ms were eliminated from the remaining data (Blocks 3, 4, 6, and 7).
3. Subjects for whom more than 10% of trials had latencies faster than 300 ms were excluded, as these subjects were likely responding without attention to accuracy.
4. Subjects were not excluded based solely on error rates. Following the recommendations of Greenwald et al. (2022), the D-score algorithm is robust to high error rates as long as subjects are attempting to respond correctly.

D-Score Computation:

1. Compute latency means (MnA1, MnA2, MnB1, MnB2) and SDs (SDA1, SDA2, SDB1, SDB2) for each of the four combined-task blocks (Blocks 3, 4, 6, and 7).
2. Compute two mean latency differences: B1–A1 = (MnB1 − MnA1) and B2–A2 = (MnB2 − MnA2).
3. Compute two inclusive (pooled) standard deviations: one using all latencies from the first pair of combined-task blocks (SD1) and another using all latencies from the second pair (SD2). [3]
4. Compute (B1–A1)/SD1 and (B2–A2)/SD2.
5. Compute the final D-score as the average of the two quotients:

$$D = \frac{(\text{B1–A1})/\text{SD1} + (\text{B2–A2})/\text{SD2}}{2} \tag{A25}$$

[3] These are computed as:

$$\text{SD1} = \sqrt{\frac{(\text{NA1}-1)*\text{SDA1}^2 + (\text{NB1}-1)*\text{SDB1}^2 + (\text{NA1}+\text{NB1})*((\text{MnA1}-\text{MnB1})^2/4)}{\text{NA1}+\text{NB1}-1}} \tag{A23}$$

$$\text{SD2} = \sqrt{\frac{(\text{NA2}-1)*\text{SDA2}^2 + (\text{NB2}-1)*\text{SDB2}^2 + (\text{NA2}+\text{NB2})*((\text{MnA2}-\text{MnB2})^2/4)}{\text{NA2}+\text{NB2}-1}} \tag{A24}$$

where *N*, Mn, and SD indicate the number of trials, means, and standard deviations for the blocks indicated (A1, B1, A2, or B2).

B.2 Intervention Materials: Feedback Template

Dear Teacher,

Thank you for participating in our research project and completing the classification test. We are writing to share your test results.

The classification test is a tool used in social psychology to measure and increase the awareness of potential preferences or unconscious associations. This test investigates the automatic associations between left-behind and non-left-behind children with positive words (e.g., "studious") and negative words (e.g., "boisterous").

- **Scenario 1: Bias against left-behind children (scores > 0.15)**
  - Score 0.15-0.35: Your classification test score is XXX, which suggests a slight automatic association between negative attributes and left-behind children. You tend to more quickly categorize positive words with non-left-behind children and negative words with left-behind children.
  - Score 0.35-0.60: Your classification test score is XXX, which suggests a moderate automatic association between negative attributes and left-behind children. You tend to more quickly categorize positive words with non-left-behind children and negative words with left-behind children.
  - Score > 0.60: Your classification test score is XXX, which suggests a strong automatic association between negative attributes and left-behind children. You tend to more quickly categorize positive words with non-left-behind children and negative words with left-behind children.
- **Scenario 2: No bias (-0.15 to 0.15)** Your classification test score is XXX, which suggests no significant automatic association between positive or negative attributes and left-behind children. You show similar response patterns when categorizing positive and negative words with both left-behind and non-left-behind children.
- **Scenario 3: Bias favoring left-behind children (scores < -0.15)**
  - Score -0.35 to -0.15: Your classification test score is XXX, which suggests a slight automatic association between positive attributes and left-behind children. You tend to more quickly categorize positive words with left-behind children and negative words with non-left-behind children.
  - Score -0.60 to -0.35: Your classification test score is XXX, which suggests a moderate automatic association between positive attributes and left-behind children. You tend to

more quickly categorize positive words with left-behind children and negative words with non-left-behind children.

- Score < -0.60: Your classification test score is XXX, which suggests a strong automatic association between positive attributes and left-behind children. You tend to more quickly categorize positive words with left-behind children and negative words with non-left-behind children.

We want to iterate that this test reveals implicit attitudes and not behaviors. Our attitudes may derive from the cultural and social context where we live, and it is not obvious that implicit attitudes and behaviors coincide. We remind you that all of your responses will be held in confidence: only the researchers involved in this study will have access to the information you provide. Your responses will not be shared with other people. Data collected will be published in aggregate form, and it will not be possible to link them with the teacher or the school. We hope that you found this test to be useful.

An enormous body of literature confirms that we all have biases—some explicit, many implicit. However, it is important to avoid our implicit biases or stereotypes related to a specific group from systematically influencing our behaviour toward students, thus influencing a child's self-image or burdening him with low expectations that will make the child feel lacking or inadequate. Acknowledging and understanding our biases can help minimize the influence they have on our daily interaction with students, including our encouragements and disciplinary procedures, track recommendations, and grades.

Thank you for the time you dedicated to our research.

The Research Team

B.3 Power Calculations

Table A2: Power Calculation

Sample size (N):	2,700 students, 100 teachers across 100 schools			
Number of clusters (k):	100 schools			
Average Cluster size ($\bar{m}$):	27 students and 1 teacher per school			
Coefficient of variation for students (CV):	0.5			
Treatment allocation (P):	50% treatment, 50% control			
Standardization (σ^2):	All outcomes standardized (mean = 0, SD = 1)			
Power ($1-\kappa$):	0.8	0.9	0.8	0.9
Significance (α):	0.1	0.05	0.1	0.05
Panel A: Teacher-level Outcomes ($N = 100$)				
Mechanism Outcomes	**Explicit bias**		**Vignette-based picture grading**	
MDE without baseline controls	0.5	0.65	0.5	0.65
Baseline correlation (ρ):	0.5 (Alwin and Krosnick, 1991)		0.5 (Karing et al., 2024)	
MDE with baseline controls	0.43	0.56	0.43	0.56
Panel B: Student-level Interaction Term ($s = 22\%$)				
Primary Outcomes	**Teacher-graded assignments**		**Teacher-Student interactions**	
Intra-Cluster correlation (ICC):	0.08		0.05	
MDE without baseline controls	0.44	0.57	0.38	0.49
Baseline correlation (ρ):	0.86		0.7 (Torsheim et al., 2000)	
MDE with baseline controls	0.22	0.29	0.27	0.35
Secondary Outcomes	**Socio-emotional outcomes**		**Blindly graded test scores**	
Intra-Cluster correlation (ICC):	0.04		0.1	
MDE without baseline controls	0.35	0.46	0.48	0.62
Baseline correlation (ρ):	0.5 (Li et al., 2025)		0.86	
MDE with baseline controls	0.30	0.40	0.24	0.32
Panel C: Sub-sample: Left-behind Student ($N = 2700 \times 0.22 = 594$)				
Primary Outcomes	**Teacher-graded assignments**		**Teacher-Student interactions**	
Intra-Cluster correlation (ICC):	0.08		0.05	
MDE without baseline controls	0.25	0.33	0.23	0.31
Baseline correlation (ρ):	0.86		0.7 (Torsheim et al., 2000)	
MDE with baseline controls	0.13	0.17	0.17	0.22
Secondary Outcomes	**Socio-emotional outcomes**		**Blindly graded test scores**	
Intra-Cluster correlation (ICC):	0.04		0.1	
MDE without baseline controls	0.23	0.30	0.26	0.34
Baseline correlation (ρ):	0.5 (Li et al., 2025)		0.86	
MDE with baseline controls	0.20	0.26	0.13	0.17

Notes: Average cluster size, coefficient of variation, ICC and baseline correlations for academic outcomes are from local education bureau records; ICC for socio-emotional outcomes and proportion of LBC are from partial baseline data (21 schools).

B.4 Measurement Instruments

B.4.1 Teacher Survey

1. Grading task Instruction: Below are paintings created by fifth-grade students. Please rate the overall quality of each work on a scale from 0 (low) to 10 (high) [Title: My Family].

Low High

0 1 2 3 4 5 6 7 8 9 10

Q6

Low High

0 1 2 3 4 5 6 7 8 9 10

Figure A2: Grading task: Pictures

To minimize potential confounding from gender stereotypes, we prepare two versions of the drawings: one depicting boys and one depicting girls. Each teacher is randomly assigned to evaluate either the boy version or the girl version. Additionally, the order in which the paintings appear is randomized across teachers to avoid order effects.

2. Explicit Bias To measure teachers' explicit attitudes toward LBC, teachers evaluate hypothetical classroom scenarios on three dimensions (teaching effectiveness, classroom discipline, and student peer relationships) using a 5-point Likert scale (1. Very poor; 2. Not very good; 3. Average; 4. Quite good; 5. Very good).

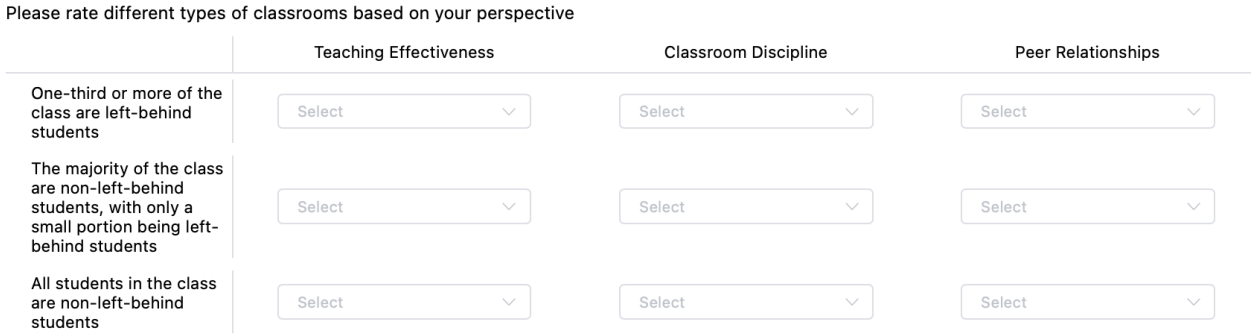


Figure A3: Explicit Bias Questions

We construct an explicit bias index by calculating the difference between teachers' ratings of classrooms with zero left-behind students (all non-left-behind) and classrooms with high proportions of left-behind students (one-third or more). We then average across the three dimensions to create an overall explicit bias score, which is standardized to have mean 0 and standard deviation 1. Higher values indicate more negative explicit attitudes toward classrooms with left-behind children.

B.4.2 Student Survey

Behavioral Problem Index Please indicate how well the following statements match your experience (1. Completely disagree; 2. Disagree; 3. Neutral; 4. Agree; 5. Completely agree):

1. I get angry when I encounter difficulties in my studies
2. I frequently argue with classmates
3. I am afraid of exams
4. I have difficulty concentrating
5. I often feel lonely
6. I am easily distracted
7. I often feel sad and upset

8. I have difficulty completing schoolwork
9. I worry that I am not performing well enough at school
10. I worry about not finishing my homework
11. I worry about not having friends to play with at school
12. I cause trouble to others by being slow and dawdling
13. I cause trouble by fighting with classmates
14. I feel ashamed when I make mistakes at school

Self-Confidence and Self-Expectations

1. Where do you think your academic performance ranks in your class? (1. Top 10%; 2. Top 25%; 3. Top 50%; 4. Bottom 50%; 5. Bottom 25%; 6. Bottom 10%)
2. What level of education do you hope to achieve? (1. Stop school now; 2. Primary school; 3. Middle school; 4. High school; 5. College; 6. Above college; 7. Indifferent)
3. How confident are you about your future? (1. No confidence at all; 2. Not very confident; 3. Quite confident; 4. Very confident)

School Sense of Belonging Think about your school. Please indicate how often the following events occur (1. Never/Rarely; 2. Sometimes; 3. Often; 4. Always):

1. If I could, I would rather not come to school
2. I feel that my classmates like me
3. I feel lonely at school

Peer Relationships

1. How many best friends do you have in your class? (fill in number)
2. Please write the names of your five best friends in the class. (If you have fewer than 5, write as many as you have)

Negative Classroom Behaviors

Participation. This semester, please indicate how many times you have done the following (1. Never; 2. Once; 3. Two to three times; 4. Once or twice a month; 5. At least once a week):

1. Said things that scared or worried other classmates
2. Hit, kicked, or pushed other students
3. Spread untrue information about other classmates on purpose
4. Purposely did not play with a classmate
5. Sent unpleasant messages or pictures to others online

Experience. This semester, please indicate how many times the following have happened to you (same scale as above):

1. A classmate said things that scared or worried you
2. Been hit, kicked, or pushed by classmates
3. A classmate spread untrue information about you on purpose
4. A classmate purposely did not play with you
5. Received unpleasant messages or pictures online

Perceived Teacher Behaviors Please indicate how often the following events occur in your class (1. Never/Rarely; 2. Sometimes; 3. Often; 4. Always):

1. My homeroom teacher notices when I am sad
2. My homeroom teacher praises me for things I do well
3. My homeroom teacher yells at me

Personality Please indicate how well the following statements match your personality (1. Completely disagree; 2. Disagree; 3. Neutral; 4. Agree; 5. Completely agree):

1. Outgoing and energetic
2. Critical and argumentative

3. Trustworthy and self-disciplined
4. Anxious and easily worried
5. Open to new things and often has new ideas
6. Introverted and quiet
7. Likeable and friendly
8. Disorganized and careless
9. Calm and emotionally stable
10. Conventional and not creative

Self-Reported Academic Outcomes

1. What was your math exam score last semester? (1. Below 60; 2. 60–70; 3. 70–80; 4. 80–90; 5. 90–100)
2. What was your Chinese exam score last semester? (same scale as above)

B.5 Index Construction: Main and Exploratory

Table A3 reports main indices and their component items. Table A4 reports exploratory indices and component analyses.

Table A3: Main Indices

Index	**Hyp.**	**Component items**	**No. items**
Explicit bias index	H2	Ratings of hypothetical classrooms on teaching effectiveness, discipline, and peer relationships; standardized average across dimensions with higher values indicating more negative attitudes toward LBC classrooms	3
Teacher behavior index	H4	Teacher notices when student is sad; teacher praises student for good performance; teacher yells at student	3
Behavioral problems index	H5	14-item behavioral problem scale from CFPS, including conduct problems (e.g., arguing, fighting), emotional difficulties (e.g., feeling sad, lonely, anxious), and peer relationship problems (e.g., difficulty making friends).	14
Well-being index (self-perception)	H5	Perceived academic standing; educational aspirations; confidence about the future	3
Well-being index (school belonging)	H5	Attendance motivation; peer acceptance; feelings of loneliness at school	3
Peer integration index	H5	Self-reported number of friends; out-degree (nominations made); in-degree (nominations received), normalized by class size	3
Peer isolation index	H5	Out-degree isolation dummy (makes no nominations); in-degree isolation dummy (receives no nominations)	2

Note: All indices are constructed as standardized averages.

Table A4: Exploratory Indices and Component Analyses

Index/Variable	Hyp.	Component items	No. items
Exploratory Indices			
Negative peer behavior — perpetration index	H5	Verbal threats; physical aggression; spreading false information; social exclusion; online harassment directed toward classmates	5
Negative peer behavior — experience index	H5	Verbal threats; physical aggression; spreading false information; social exclusion; online harassment experienced from classmates	5
Component Analyses (individual items)			
Perceived teacher — notices emotions	H4	Teacher notices when student is sad	1
Perceived teacher — praises students	H4	Teacher praises student for good performance	1
Perceived teacher — negative interactions	H4	Teacher yells at student	1
Self-perceived ability	H5	Perceived academic standing	1
Self-confidence	H5	Confidence about the future	1
Self-expectations	H5	Educational aspirations	1
Attendance motivation	H5	Desire to attend school	1
Peer acceptance	H5	Perceived likability among classmates	1
Loneliness at school	H5	Feelings of loneliness at school	1
Peer isolation — out-degree	H5	Binary indicator: student makes no friendship nominations	1
Peer isolation — in-degree	H5	Binary indicator: student receives no friendship nominations	1
Self-reported friends	H5	Number of self-reported friends in classroom	1
Out-degree	H5	Number of friendship nominations made, normalized by class size	1
In-degree	H5	Number of friendship nominations received, normalized by class size	1

Note: Exploratory indices and component analyses complement the main indices. We may also explore alternative indices in exploratory analyses such as the GLS-weighted summary index Anderson (2008).