Learning with Peers, at Scale: Experimental Evidence from an Embedded Peer Tutoring Intervention*

Palaash Bhargava †

Madhavi Jha ‡

Dashleen Kaur §

Nikhil Kumar¶

Tarang Tripathi‖

Do not circulate this version.
September 17, 2026

We study a light-touch peer tutoring intervention (at scale) facilitated by teachers and embedded within rural public primary schools in the poorest Indian state of Bihar in partnership with the district education administration and an NGO. Teachers in treatment schools were encouraged to form and maintain groups of size 5-6 with a high-performing student as the tutor and conduct peer tutoring sessions at least thrice a week in one period of the regular school day. We find that the program increased math performance by 0.15-0.16 standard deviations (s.d.) for both learners and tutors driven by gains in learning across a range of math competencies. Learners in treatment schools report finding studies easier (+0.11 s.d.) and feeling less anxious when studying with friends (-0.1 s.d.). We do not find any effects either for tutors or learners on a range of socioemotional skills. We also find a tightening of student peer groups such that both tutors and learners are reported by a smaller number of peers as part of study groups (-0.14 for learners and -0.27 for tutors) while learners also report studying with fewer peers (-0.20). Both learners as well as tutors are reported as part of better study groups. We do not find any effect on the size or quality of friendship networks. These results demonstrate that same-grade peer tutoring can be implemented by teachers at scale within public primary school classrooms effectively.

*We acknowledge with gratitude the support of The Agency Fund. We are especially grateful to Emily Breza, Sandra Black, Rema Hanna, Gabriel Kreindler, Eliana La Ferrara, Matt Lowe and Nathan Nunn for their invaluable feedback. We are also grateful to Augustin Bergeron, Michela Carlana, Andreas de Barros, Will Dobbie, Asim Khwaja, Amanda Pallais, Heather Sarsons and Kartik Srivastava for helpful comments. We thank the participants of the Development, Labor, and Economic and Social Policy workshops at Harvard University as well as participants at NEUDC. We are grateful to Involve Learning Solutions Foundation for supporting this study and thank Anna Daniel, Ashwini Ramakrishnan, Dr. Pratibha Narayanan and Ritupma Shekhawat for their excellent project leadership and field coordination. We thank Uddyalok Bangabash and Chandraditya Raj of Aawaz Foundation for their excellent support in the data collection process. Data protection was secured in compliance with the relevant local legal provisions and regulations concerning Human Subjects research at Columbia University, Harvard University, University of British Columbia and University of California San Diego. The project has ethics approvals from the Committee on the Use of Human Subjects at Harvard University (protocol IRB24-1240). The trial has been registered at the AEA Registry: AEARCTR-0015527.

†University of Chicago. E-mail: palaash.bhargava@columbia.edu

‡Harvard University. E-mail: madhavijha@g.harvard.edu

§University of British Columbia. E-mail: dashleen@student.ubc.ca

¶Harvard University. E-mail: nikhil_kumar1@g.harvard.edu

‖University of California San Diego. E-mail: tatripat@ucsd.edu

1 Introduction

Over the past 25 years, primary school enrollment has increased dramatically, driven in part by large public investments and international commitments. However, this has not translated into gains in learning. This is true in many low- and middle-income countries (Bank, 2022). In India, for instance, nearly half of students completing grade 5 cannot solve basic arithmetic problems expected at grade 2, and almost half cannot read a grade 2-level text (ASER, 2025). This disconnect between access and quality represents a critical barrier to the accumulation of human capital in these settings.

The core problem is not simply a lack of schools or teachers, but how instruction is organized within classrooms. Pedagogy in resource-constrained settings remains predominantly teacher-centric, focused on syllabus completion, rote learning, and recitation. Students have limited opportunities for active learning or peer interaction and there is hardly any differentiated instruction tailored to students' varied preparation levels (Glewwe and Muralidharan, 2016). India's 2020 National Education Policy proposes peer tutoring as one way to improve the classroom environment in schools calling on states to adopt "innovative models to foster peer tutoring under the supervision of trained teachers" (GoI, 2020). Can peer tutoring, a low-cost intervention requiring no new resources, improve learning at scale when implemented by regular teachers in government-run primary schools?

Peer tutoring may provide learners with more individualized, accessible explanations from peers who may be closer to their level of understanding and better positioned to identify conceptual difficulties. High-achieving peers have been shown to improve the learning of their low-achieving counterparts (Li et al., 2014; Fafchamps and Mo, 2018). It may also benefit tutors through the "learning by teaching" as explaining concepts to others requires cognitive reorganization, deeper engagement with material, and metacognitive reflection, all of which consolidate understanding (Shah, 2025; Freeman et al., 2014). Beyond cognitive gains, peer tutoring may foster socioemotional development by creating structured opportunities for collaboration, responsibility, and empathy (Alan and Ertac, 2018) - skills increasingly recognized as critical determinants of long-term socioeconomic outcomes (Deming, 2017; Chetty et al., 2022). Importantly, unlike resource-intensive school reforms as well as other forms of tutoring, peer tutoring requires minimal financial input and can be implemented entirely within existing school structures, by changing how classrooms within schools are organized and relying on teachers to facilitate rather than on external tutors, supplemental materials, or technology.

Yet, peer tutoring could harm learners if same-grade tutors lack the pedagogical skills or content mastery to explain concepts as effectively as teachers. It could also slow tutors' own learning if tutoring diverts time and cognitive effort away from their academic development. Additionally, formal tutor assignments might create unintended social dynamics - learners excluded from tutor roles may experience diminished status, while designated tutors face elevated social pressure or expectations that could affect classroom relationships. Hence, whether same-grade peer tutoring

delivers net benefits when implemented by regular teachers with minimum support at scale in government schools remains an open empirical question.

While tutoring interventions have been shown to broadly generate substantial learning gains in small-scale studies (Nickow et al., 2020; Fryer, 2017), meta-analyses of tutoring programs demonstrate that effect sizes decline sharply as interventions scale (Kraft et al., 2024). This raises questions about whether peer tutoring, implemented by regular teachers rather than external tutors, can maintain effectiveness in large-scale government systems. Moreover, most peer tutoring evidence comes from cross-grade programs (older students tutoring younger ones as in Romero et al. (2022)) or from highly controlled interventions with limited external validity (Gal and Fallik, 2021).

We address these evidence gaps by studying a large-scale, teacher-led same-grade peer tutoring intervention in India's poorest state of Bihar in the district of Bhagalpur. The district is characterized by high student-teacher ratios, limited infrastructure, widespread multi-grade classrooms, and severe foundational learning deficits. Bhagalpur provides an ideal setting to examine both the scalability and robustness of the effects of peer tutoring under conditions of limited resources, heterogeneous student learning levels and traditional classroom environments.

In partnership with Bhagalpur District Institute of Education and Training (DIET) and Involve Learning Solutions Foundation (Involve), an NGO, we implemented a light touch peer tutoring intervention with the help of existing school teachers in government-run primary schools in grades 3 through 5. Involve designed the program to complement teacher instruction. Specifically, at the start of the academic year, teachers in all the study schools were informed about peer tutoring as part of their annual training while only the teachers in treatment schools received training in how to implement peer tutoring in classrooms. Teachers in treatment schools were also provided with suggested peer group assignments based on baseline assessments, and were supported by monthly monitoring visits from Involve. Teachers in treatment schools were encouraged to implement peer tutoring sessions at least three times a day with a 40-minute period reserved for peer tutoring sessions on these days. Teachers informed tutors about their responsibilities and supervised peer tutoring implementation according to their training at the start of the study.

To form peer groups, we ranked students in each grade in a school based on a baseline mathematics assessment for fundamental numeracy, and those in the top 20 percentile in each class were designated as peer tutors. Each tutor was randomly assigned a group of four peer learners from the same grade. Crucially, teachers retained discretion to modify group assignments based on their knowledge of students - a design feature that mirrors real-world implementation conditions and allows us to use teacher agency for program delivery at scale (List, 2022). In fact, we observe that while teachers maintained most of the peer tutors' status within the peer tutoring groups, teachers regularly changed the learners in these groups.

We evaluate the intervention through a stratified cluster randomized controlled trial involving 177 schools and approximately 14,000 students in grades 3 through 5. We collect rich data on mathematics learning achievement, classroom experiences, socioemotional skills, and social networks at baseline and endline as well as information on demographic and household characteristics, allowing us to examine not only whether peer tutoring improved learning, but also how it reshaped classroom environment, social networks and socioemotional skills.

Our results show five key findings. First, peer tutoring substantially improved learning in mathematics: students in treatment schools scored 0.16 standard deviations (s.d.) higher than those in control schools, driven by improvements in medium difficulty (+0.09 s.d.) and high difficulty (+0.13 s.d.) questions (classified based on baseline responses) without any effect on the easiest questions. This indicates that peer tutoring improved learning across different types of competencies. Second, we find similar effects for learners (+0.16 s.d. over a control mean of -0.043 s.d.) and for tutors (0.15 s.d. over a control mean of 0.50 s.d.). This suggests that peer tutoring improved the learning of not just the low-achieving students by the high-achieving student tutors as well, suggesting benefits from "learning by teaching".

Third, the program modestly learners ease of learning (+0.11 s.d.) and reduced anxiety when studying in groups (-0.1 s.d.) while the effects for tutors are not statistically significant but in the same direction. Fourth, we find no statistically significant effects on a range of socioemotional skills. Lastly, we find that study groups shifted in structure, with both tutors and learners reporting smaller study groups without any changes in their size of friendship or help-seeking networks. Moreover, we also find that learners become part of higher performing study groups relative to their counterparts in the control group. These results show that peer tutoring can lead to improvements across the distribution of learning levels, demonstrating that peer tutoring is not limited to helping a narrow subset of students but raises learning levels broadly within classrooms. Taken together, our findings suggest that peer tutoring is an effective, scalable, and institutionally feasible complement to teacher instruction, all without additional personnel, infrastructure, or financial inputs beyond light-touch teacher training and monitoring.

We contribute to several strands of literature. First, we provide rare experimental evidence that same-grade peer tutoring can improve learning at scale in public primary schools in a low-resource setting, extending findings from cross-age and secondary school contexts (Romero et al., 2022; Li et al., 2014; Fafchamps and Mo, 2018) to younger learners at foundational stages of education. Second, we document clear evidence of "learning by teaching" (Shah, 2025; Berlinski and Busso, 2017) by showing that tutors gained as much as learners, demonstrating that peer tutoring benefits students across the achievement distribution rather than merely redistributing learning from high to low achievers. Third, we show that peer tutoring can be implemented by regular teachers without requiring wholesale pedagogical reform, addressing a key scalability constraint identified in prior research on active learning and school reforms (Freeman et al., 2014; Berlinski and Busso, 2017; Bold et al., 2017). Fourth, we contribute to the literature on scaling educational interventions in general and tutoring interventions in particular,

by demonstrating peer tutoring's effectiveness in a real-world implementation involving a large number of government-run schools (List, 2022; Kraft et al., 2024).

The rest of the paper proceeds as follows. In section 2, we describe the context of our intervention. In section 3, we present the details of the intervention and experimental design. In section 4, we describe the data that we collect and present some motivating facts. In section 5, we present the empirical strategy and discuss the results from the experiment. We conclude in section 6.

2 Context

The Right of Children to Free and Compulsory Education (RTE) Act, legislated in 2010, establishes that every child in India has "a right to full time elementary education of satisfactory and equitable quality in a formal school which satisfies certain essential norms and standards" (GoI, 2010). India's National Education Policy, 2020 (GoI, 2020) envisions the curricular and pedagogical structure for school education as consisting of the Foundational Stage (Anganwadi/pre-school in age 3-5 and lower primary school grades 1-2 in age 6-8), Preparatory Stage (primary school grades 3-5 in ages 8-11), Middle Stage (upper primary school grades 6-8 in ages 11-14), and Secondary Stage (lower secondary school grades 9-10 in ages 14-16 and upper secondary school grades 11-12 in ages 16-18). In this study, we focus on the preparatory stage i.e. primary school grades 3-5.

Almost three decades of national public primary education programs (District Primary Education Programme (DPEP), Education for All Campaign (Sarva Shiksha Abhiyan, SSA), Right to Education Act (RTE)) have expanded access to primary schools in India. Since 2001, these programs has brought more than 20 million children into primary schools. In 2023-24, government-run or -aided primary schools were the most common type of schools that provided primary education in India, accounting for about 71 % of schools providing primary education in the country.

Our intervention was conducted in Bhagalpur district in Bihar which is the second-largest Indian state in terms of population. In Bihar, the enrollment rates for students aged 6-14 are very similar to the rest of the country while the learning rates are similarly poor. This can be observed in table 1 which also shows the percentage of students in Bihar enrolled in a grade who demonstrative the respective skill. Clearly, a large fraction of students exiting grade 5 do not learn basic literacy and numeracy skills that they should have mastered in the previous few years of schooling.

At the same time, Bihars schools face acute infrastructural and institutional challenges. While access to basic provisions such as mid-day meals (92.9 %) and drinking water (88.76 %) is high, fewer schools provide libraries (67.9 % versus 82.5 % nationally), playgrounds (47.9 % versus 68.9 % nationally), or computers (16.5 % versus 27.4 % nationally). Teacher presence is also lower:

Table 1: ***Student Learning Levels by Skill and Grade (ASER 2024)***

Region	Grade 2 Math		Grade 3 Math		Grade 2 Reading	
	Grade 3	Grade 5	Grade 3	Grade 5	Grade 3	Grade 5
India	33.7 %	55.8 %	11.4 %	30.7 %	27.0 %	48.7 %
Bihar	37.4 %	55.7 %	19.5 %	36.0 %	26.1 %	43.6 %

on average, only 79 % of teachers are present in Bihar schools during ASER visits, compared to 87.5 % nationally (see table 2). These deficits constrain the learning environment, particularly in government-run schools, which serve the majority of students in the state.

Table 2: ***School Facilities and Educational Inputs(ASER 2024)***

Item Description	India	Bihar
School facilities		
Mid-day meal served in school on day of visit	91.9	92.9
Drinking water available	77.7	88.7
Boundary Wall	75.3	59.3
Toilet usable	79.0	82.5
Girls toilet: Separate provision, unlocked and useable	72.0	73.6
Library	82.5	67.9
Playground	68.9	47.9
Schools with electricity available on day of visit	89.7	91.6
Computer available for children to use	27.4	16.5
Educational Inputs		
Enrolment: Age 6-14	98.1	97
% Enrolled children present	75.9	60.9
% Teachers present	87.5	79
School has Multi-grade classrooms	66	72
School complies with pupil-teacher ratio norms	63.5	38.2

In the district of Bhagalpur specifically, 95 % of children aged 6-14 are enrolled in a school while 77.5 % of children in this age-group are enrolled in a government-run school. 55.1 % of students enrolled in grade 3-5 can solve a grade 2 level math problem while only 44.1% of students enrolled in grades 3-5 can read a grade 2 level text.

These deficiencies are not new. Between 2012 and 2022, learning levels in Bihar stagnated, with nearly identical gaps recorded in by earlier rounds of Annual State of Education Reports (ASER, 2025). Studies of classroom practices point to systemic drivers of this crisis: instruction in Bihars schools is largely textbook-driven, with little adaptation to the actual learning levels of students (Sinha et al., 2016). Teaching is predominated by rote learning, with students seated in rows and working individually, and multigrade classrooms remain widespread: 72 percent of schools in Bihar report combining classes across grades 1-5 (ASER, 2025). This combination of heterogeneity in childrens initial preparation levels within classrooms, focus on syllabus completion, rigid

pedagogy and large student-teacher ratios has compromised foundational learning in primary school grades.

To address the learning crisis and improve Foundational Literacy and Numeracy (FLN) skills, the National Education Policy (2020) recommends that "peer tutoring can be taken up ... under the supervision of trained teachers" and exhorts states to establish "innovative models to foster such peer-tutoring". This paper studies the effect of one such light-touch model of peer-tutoring.

Against this backdrop, the National Education Policy (2020) emphasizes the importance of Foundational Literacy and Numeracy (FLN) and specifically recommends that states adopt "innovative models to foster peer tutoring under the supervision of trained teachers." Our study directly evaluates such a model, institutionally embedded in the preparatory stage of school (grades 3-5). By situating our intervention in Bhagalpur, a district characterized by high poverty, widespread first-generation learners, multi-grade classrooms, and chronic learning gaps, we test not only whether peer tutoring can raise student achievement, but also whether it can do so under the conditions of real-world classrooms of low-resource settings in which such interventions must ultimately operate.

3 Intervention details and Experimental design

The intervention was implemented in partnership with Bhagalpur District Institute of Education and Training (DIET), a government body whose objective is to train teachers within the district, as well as Involve Learning Solutions Foundation (Involve), a non-profit organization in India that has been developing peer tutoring modules for primary schools since 2016. It is grounded in the idea that empowering students as tutors can both improve classroom learning and strengthen social networks.

Prior pilots run by Involve around Bengaluru (Karnataka, a state in south India) demonstrated feasibility, and the interventions was adapted for large-scale implementation in government-run schools in Bihar. Specfically, while Involve directly implemented the peer tutoring intervention in the pilot schools in Bengaluru, in order to scale up their program to an entire district, Involve worked with DIET Bhagalpur to embed training of teachers to implement the peer tutoring intervention in their annual teacher training.

In each grade in a school, students were ranked based on their scores on a baseline mathematics skills assessment. Students who were in the top 20% of their grades were assigned as peer tutors and the rest became peer learners. Each tutor was grouped with a randomly selected group of 4 peer learners within their grade. To reinforce the structure, visual charts (see Appendix figure A1) listing the tutors and their groups were displayed in treatment classrooms. Peer tutors received guidance from teachers to understand their responsibilities. In treatment schools, one 40-minute period was reserved at least thrice a week for peer tutoring sessions, where tutors

worked with their groups on mathematics skills (See Appendix figure A2 for an example of peer tutoring in progress).

An important feature of the design was that the top 20% of the students were selected as tutors from each school. Since we have a similar counterfactual students for control schools, we can compare outcomes for students who were assigned as tutors (in treatment schools) to those who were eligible but not selected (since they were in control schools), providing an additional layer of identification for estimating the effects of being a tutor itself. To ensure implementation fidelity, Involves field team conducted regular school visits, monitored sessions, and supported teachers.

3.1 Evaluation Design

DIET Bhagalpur provided 6 blocks (of 16) for the purpose of this study, which had a total of 488 government-run primary schools according to the publicly available data from Unified District Information System for Education (UDISE), a database developed at the Department of School Education, Ministry of Education, Government of India [1]. To sample schools into the study and to randomly assign schools to treatment, we proceed in the following two steps.

3.1.1 Sampling of schools

First, within each block, we first restricted ourselves to the schools between 50th and 95th percentile of total enrolment in grades 3 to 5 in that block. This yielded 232 schools. Then, within each block, we randomly sampled schools proportional to the share of schools in that block in the 232 schools with the goal that we end up with 178 schools. Table 3 describes the differences between schools included in the study sample and those that are not part of the study sample. Schools in the study sample have more teachers, a larger student-teacher ratio as well as more students in grades 3-5.

3.1.2 Randomization

To randomly assign schools to treatment and control groups, we used stratified randomization. Specifically, we constructed principal components based on (i) the number of teachers, (ii) the number of students in grades 3 to 5, and (iii) distance to Bhagalpur city. We then split the schools above and below the median of the first two principal components, creating four strata within each block. Inside every block and strata, we then randomly assigned the schools to treatment and ended up with 82 schools in the treatmenta arm and 96 in the control arm. Table 3 also shows the difference in the characteristics of the schools randomly assigned to treatment and control arms. We do not observe any statistically significant differences.

[1] See https://udiseplus.gov.in//en/home

Table 3: *Sample of schools, representativeness, and balance tests*

	Representativeness (overall)			Balance Checks		
	Non-Study	Study	Difference	Control	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Number of teachers	5.38	7.31	1.93***	7.34	7.28	-0.08
	[3.05]	[2.84]	(0.28)	[2.89]	[2.80]	(0.30)
Dist to Bhagalpur (in km)	16.83	16.27	-0.55	16.29	16.24	-0.34
	[9.06]	[8.81]	(0.37)	[8.66]	[9.04]	(0.35)
Student Teacher Ratio	16.64	20.67	4.06***	20.83	20.48	-0.18
	[10.29]	[9.74]	(0.93)	[10.87]	[8.29]	(1.28)
Number of students in 3-5	85.21	134.76	49.71***	135.85	133.48	-1.45
	[70.51]	[41.81]	(5.67)	[44.24]	[39.01]	(3.98)
Fraction of Female teachers	0.46	0.43	-0.03	0.45	0.41	-0.04
	[0.27]	[0.21]	(0.02)	[0.21]	[0.21]	(0.03)
Fraction of General teachers	0.24	0.24	-0.00	0.25	0.22	-0.03
	[0.23]	[0.21]	(0.02)	[0.22]	[0.20]	(0.03)
Number of schools	310	178	488	96	82	178
Joint F-test (p-value)			0.00			0.61

Notes. This table reports on the study's sample of schools. Study refers to the 178 public primary schools included in the study's effective sample. Non-study refers to all other public primary schools in Bhagalpur. Control refers to the 95 schools not assigned to receive the peer tutoring intervention, and Treatment refers to 82 schools assigned to receive this intervention. Difference reports on the regression-adjusted difference. Standard deviations are shown in brackets; standard errors are shown in parentheses. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find that some student characteristics were unbalanced but a joint F-test does not reject the null of no difference (see tables A1 and A2).

3.2 Implementation

Teachers from all teachers received information about peer tutoring and its benefits as part of their training at the start of the school year. Due to administrative constraints, a pure control group that does not receive any information about peer tutoring could not be established. However, only treatment schools are assisted by our implementing partner, Involve, in peer-tutoring. Involve tried to ensure that peer tutors and peer groups are assigned according to the original program design, facilitated by their field team through regular school visits to oversee the implementation. Hence, we interpret the intent-to-treat effects we report below as the lower bound as some peer tutoring may have happened in the control schools.

Baseline surveys were conducted in July 2024, prior to the rollout of the intervention. The intervention was then launched immediately after the baseline and continued throughout the 2024-25 school year. Teachers from treatment schools were to implement peer tutoring in August

2024 and were later provided the groups formed based on the baseline assessments. In treatment schools, daily peer tutoring sessions were implemented and monitored, with regular visits by Involve field team to provide support. Endline surveys were conducted in February-March 2025, at the conclusion of the academic year (see timeline in Appendix figure A3).

4 Data and Descriptive Statistics

The sample for this study consisted of all students in grades 3, 4 and 5 in the 177 study schools. All surveys were conducted in person inside the classrooms in the schools where students are enrolled in.

4.1 Data

We use a comprehensive set of indicators to characterize the mathematical skills, socioemotional well-being as well as students' social networks. The main data used in this paper comes from baseline and endline surveys conducted in classrooms of the study schools. This included all students in grades 3, 4 and 5 in the study schools. In the survey, students were asked questions about five different topics: 1) mathematical skills 2) classroom and learning experience 3) socio-emotional skills 4) social networks 5) socio-economic status 6) peer learning experience (only at endline). These are described in detail below.

1. **Mathematical Skills** The assessment included questions about comparing numbers, counting numbers, visual comparison of length, addition, subtraction, multiplication, subtraction, pattern recognition and related word problems.

2. **Classroom and learning experience**: We asked students about how easy they find learning, their fear of Math, their confidence in participating in classrooms etc.

3. **Socioemotional well-being**: We measured aspiration, leadership, competitiveness, confidence, grit, growth mindset, patience, pro-sociality, and the big 5 traits (agreeableness, conscientiousness, extroversion, neuroticism, openness).

4. **Social Networks**: To map classroom networks, we asked students to report names and grades of their friends, of students who they seek help from in studies and of students who have the most friends.

5. **Socioeconomic status**: We also asked questions about student's father's and mother's occupation and ownership of assets like cycle, motorbike, car, television, computer etc. to measure the students' socio-economic status.

6. **Peer learning experience (endline only)**: To measure students' peer learning experience, we included questions about students' peer learning group, tutor status, experience of peer learning, frequency of peer learning sessions etc. These questions were only asked in the endline survey.

4.2 Baseline Descriptive Statistics

Using data from the baseline survey of 14,056 students across our 177 study schools, we describe student learning levels, their classroom experiences, socioemotional skills and social networks.

In figure 1, we plot a probability distribution of the student learning levels at baseline by grade. We observe that the distribution shifts to the right in higher grades which is evidence of learning as students progress across grades (also see figure A4). At the same, we also observe a substantial share of students in grade 5 who are at the same learning level as those in grade 4 or grade 3.

Figure 1: *Distribution of Student Learning Levels at Baseline by Grade*

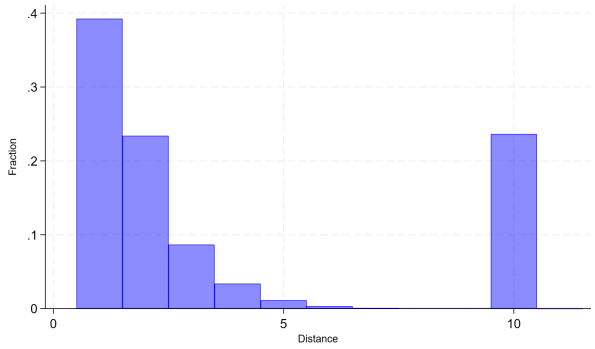


In figure A5, we show the distribution (in the form of boxplots) of the fraction of students in each grade in a school that belong to each decile of the baseline learning levels and observe that schools vary considerably in their share of students at different learning levels in each grade. For example, in grade 3, the percentage of students in a school that belong to the lowest decile ranges from 2 % to 80 %. Taken together, these suggest a large heterogeneity in learning both within and across schools for students enrolled in the same grade level.

In figure 2, we plot a probability distribution of the distance from the top 20 percent students in the grade (based on baseline scores) of the rest of the students in the same grade in a school and find that more than half the students are not directly connected to the high-performing students. Specifically, less than 40 percent students report seeking help from a high-performing students and less than 50 percent students report being friends with one of them.

Figure 2: *Distribution of social distance to the top 20 percentile students in the same grade*

(a) *: Help-seeking network*



(b) *Friendship network*

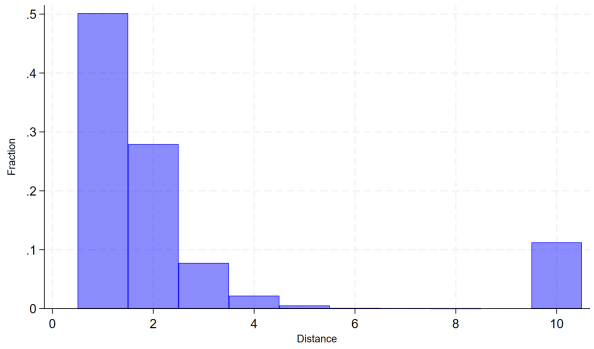
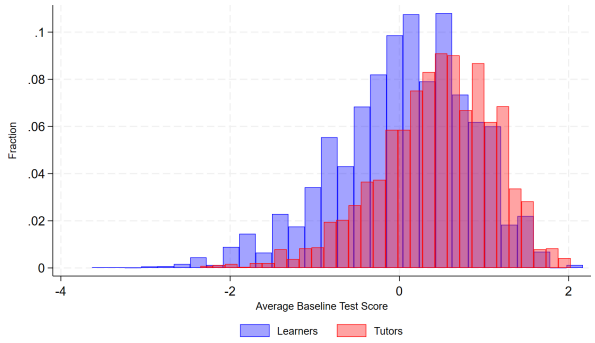
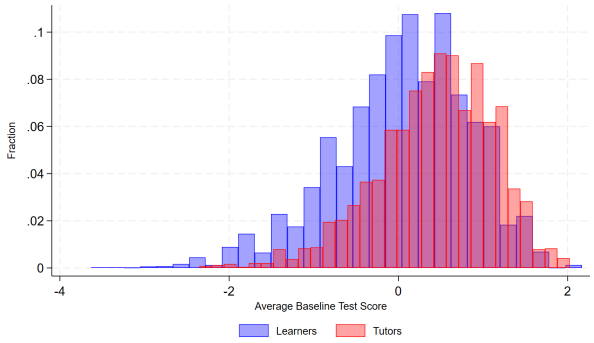


Figure 3: *Distribution of average scores of "incoming links"*

(a) *Help-seeking networks*



(b) *Friendship networks*

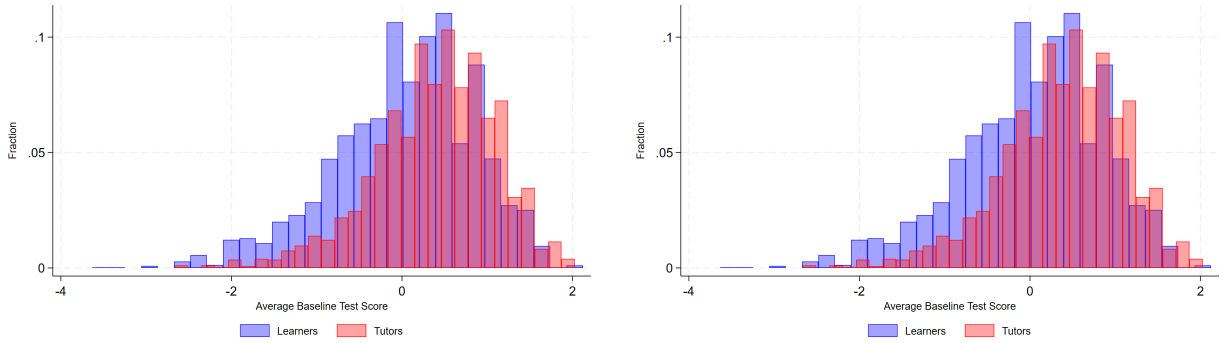


In figure 3, we show the distribution of the average baseline score of peers who report a student as part of their network, separately for high-achieving students and the rest. We find that low-performing students are sought for help in studies or named as friends by peers with lower average baseline scores than high-performing students. Conversely, in figure 4, we show the distribution of the average baseline score of peers who a student reports as part of their network, separately for high-achieving students and the rest. We find that low-performing students seek help in studies for or name as friends peers with lower average baseline scores than high-performing students. In appendix figures A6- A9, we show similar distributions dissagregated by grade and find similar patterns across grades. Taken together, these suggest endogenously positive assortative mixing in the classroom at baseline.

In light of these facts, the peer tutoring intervention we study should be seen as shifting the entire distribution in figure 2 to the left and shifting the social networks such that the average scores of the network of learners and tutors (figures 3 and 4 become closer.

Figure 4: *Distribution of average scores of "outgoing links"*

(a) *Help-seeking networks* **(b)** *Friendship networks*



In table 4, we present the pair-wise correlation between math scores and socioemotional skills at baseline. We find that students who form better at baseline are less shy in asking questions in class, find studying easier, are less afraid of math, enjoy homework more, place more importance on going better in studies, are more likely to believe that hardwork improves grades, aspire more to be a leader, are more gritty and are less prosocial.

Table 4: *Baseline: Correlation between socioemotional traits, test scores and assigned tutor status*

Variable	Test Score	Leader
Confident Asking Qs	0.152***	0.222***
Studying is easy	0.149***	0.218***
Afraid of maths	-0.189***	-0.266***
Scared of HW	-0.188***	-0.254***
Competitiveness	0.202***	0.254***
Growth Mindset	0.196***	0.248***
Aspiration to Become Monitor	0.193***	0.259***
Grit	0.146***	0.224***
Prosociality	-0.029**	-0.050**
Agreeableness	-0.091***	-0.141***
Extraversion	0.127***	0.171***

In table 5, we present the pair-wise correlation between math scores and the size of student social networks at baseline. We find that students who perform better at baseline also have larger social networks.

Table 5: *Baseline: Correlation between network characteristics, test scores and assigned tutor status*

Variable	Test Score	Tutor
# students helped	0.476***	0.840***
Average score of students helped	0.425***	0.721***
# students seek help from	0.249***	0.307***
Average score of students seek help from	0.431***	0.636***
# friendship nominations received	0.458***	0.699***
Average score of friendship nominations received	0.627***	0.956***
# friends nominated	0.542***	0.770***
Average score of friends nominated	0.602***	0.939***
# nominations for most friends	0.354***	0.601***

4.3 Groups Assignment and Implementation

Teachers did not follow the prescribed group assignment exactly. Teachers systematically chose tutors with lower learning levels (as measured at baseline) and created peer tutoring groups with larger and more varied learning differences between tutors and learners (than the assigned groups), while preserving the learning distribution among learners themselves. We provided more details on this below.

4.3.1 Changes to assigned tutors

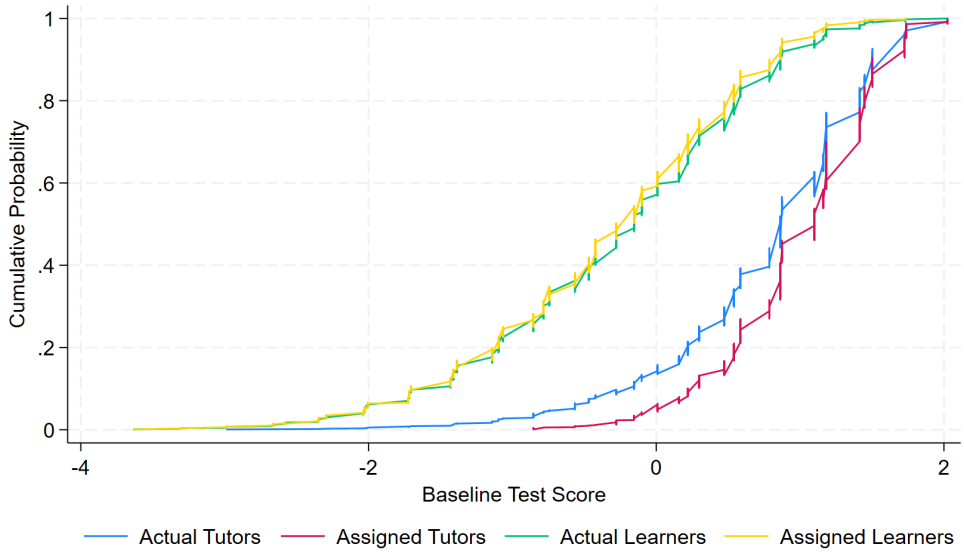
Figure 5 shows the cumulative distribution function of learning scores for assigned and actual tutors and learners. While the assigned tutors were at the extreme right of the distribution, the actual tutors that teachers chose during the implementation of peer tutoring were slight to the the left of the assigned tutors, suggesting that teacher likely considers non-academic factors into account when choosing peer tutors.

We also see that the actual learners and assigned leaders are not very different from each other except to the extent that teachers chose some assigned tutors to become learners and some assigned learners to become tutors.

4.3.2 Changes to assigned groups

In figure A10, we observe that teachers shuffled the groups such that the distribution of average distance between tutors and learners in implemented groups had a larger mean and a larger variance than those in assigned groups. By increasing the average tutor-learner distance (relative to random group formation), teachers likely thought that pairing stronger tutors with weaker learners would maximize learning gains for struggling students. But the increased variance suggests teachers didn't apply a blanket rule, but instead made context-specific judgments (e.g.,

Figure 5: *Tutors and Learners: Assignment and Implementation*



"Tutor A can handle tutoring peers with very low ability, but Tutor B would be intimidated by too large a gap") or balanced learning gaps with other factors like personality or social dynamics.

Moreover, in figures A11, A12 and A13 respectively, we observe that the distribution of standard deviation of the distance between tutors and learners, the distribution of average distance between learners and the distribution of standard deviation of the distance between learners was similar between assigned and implemented groups.

Taken together, these patterns suggest that teachers systematically deviated from assigned groups to form groups with larger and more varied learning differences between tutors and learners, while preserving the learning composition among learners themselves. These adjustments provide insight into teachers' implicit theories about optimal peer tutoring groups which we leave for future research to explore.

5 Empirical Strategy and Results

5.1 Empirical Model

We estimate the intent to treat (ITT) effect of the program on outcomes of interest by conditioning on baseline covariates, baseline outcome and randomization strata-by-grade fixed effects.

5.1.1 Aggregate Effects

In our main specification, we estimate the following model:

$$Y_{igsr} = \alpha + \beta T_s + \delta Y_{0igsr} + \Pi X_{igsr} + \gamma_{gr} + \varepsilon_s \tag{1}$$

where Y_{igsr} is the outcome for student i in grade g in school s in randomization strata r at the endline, T_s is an indicator for whether the school was randomized to receive the treatment, β is our coefficient of interest, the ITT estimate, Y_{0igsr} is the outcome for student i in grade g in school s in randomization strata r at baseline, X_{igsr} are characteristics of student i in grade g in school s in randomization strata r namely age, gender and socioeconomic status and γ_{gr} are randomization strata-by-grade fixed effects. We cluster standard errors at the school level.

Panel A of table 6, A3, 7, 8, 9 and 12 report this β, the intent-to-treat (ITT) effects of the peer tutoring program on the fundamental numeracy skills of all students, on average.

5.1.2 Effects by Tutor Status

We then disaggregate this average ITT effect by whether the student was assigned as a tutor in their school by interacting the treatment status of the school with the tutor status of the student. Specifically, we estimate the following model:

$$Y_{igsr} = \alpha_1 + \beta_1 T_s + \beta_2 \text{Tutor}_{igsr} + \beta_3 T_s \times \text{Tutor}_{igsr} + \delta Y_{0igsr} + \Pi X_{igsr} + \gamma_{gr} + \varepsilon_s \tag{2}$$

where Y_{igsr} is the outcome for student i in grade g in school s in randomization strata r at the endline, T_s is an indicator for whether the school was randomized to receive the treatment, Tutor_{igsr} is the tutor status of the student, Y_{0igsr} is the outcome for student i in grade g in school s in randomization strata r at baseline, X_{igsr} are characteristics of student i in grade g in school s in randomization strata r namely age, gender and socioeconomic status and γ_{gr} are randomization strata-by-grade fixed effects. We again cluster standard errors at the school level.

In this specification, β_1 is the ITT estimate for the effect of the program on learners, β_2 is the difference between tutors and learners in schools in the control group, β_3 is the coefficient on the interaction term $T_s \times \text{Tutor}_{igsr}$ such that $\beta_1 + \beta_3$ is the ITT estimate of the effect of the program on tutors and $\beta_2 + \beta_3$ is the difference between tutors and learners in treatment schools.

Columns 3-8 of tables 6, A3, 7, 8, 9 and 12 present the ITT effects disaggregated by whether the student was assigned to be a tutor or not. Column 3 reports the average outcome for learners in control schools, column 4 reports the ITT effect on outcomes for learners due to the program (β_1), column 5 reports the average difference in outcomes between tutors and learners in control schools (β_2), column 7 reports the ITT effect on outcomes for tutors due to the program ($\beta_1 + \beta_3$), column 8 reports the average difference between tutors and learners in

treatment schools ($\beta_2 + \beta_3$), and column 6 reports the the difference between the treatment and control schools in the outcome gap between tutors and learners (β_3).

5.2 Results

5.2.1 Effects on Academic Achievement

Based on our main empirical specification 1, in panel A of table 6, we find that the program led to a 0.16 standard deviations (s.d.) increase in the scores of students on a test of their math ability (column 1). This is robust to weighting the questions by their level of difficulty (column 2). Moreover, in columns 3-5, we find that the effect comes primarily from students performing better in medium and high difficulty questions (classified by students' performance on each question in the baseline). Specifically, we find an effect of 0.09 s.d. on medium difficulty questions and an effect of 0.13 s.d. on high difficulty questions as well as a statistically insignificant effect of 0.06 s.d. on the easiest questions. In panel A of table A3, we disaggregate the effects into different types of competencies.

Based on specification 2 that disaggregates the ITT effects by the assigned tutor status of a student, we find in panels B and C of table 6 that the program led to a 0.16 s.d. increase in scores of learners and 0.15 s.d. increase in scores of tutors. Since this difference is not statistically significant, it suggests that the program improved the fundamental numeracy skills of learners as well as tutors. It is important to note that the improvement for tutors was over a control mean of 0.496 s.d. while that for learners was over a control mean of -0.043 s.d.. We also find that both tutors and learners benefited similarly also when disaggregating the score by level of difficulty.

We also find broad-based gains across the distribution of baseline learning levels. Figure 6 illustrates this non-parametrically, and plots the average student test-scores at the endline against their within-grade baseline percentiles, separately for treatment and control groups. Across most baseline percentiles and for all three grades, the averages shifts upward for the treatment group, indicating broad-based gains across the learning distribution.

5.2.2 Effects on Learning Experience

Based on our main empirical specification 1, panel A of table 7 shows that on average, the program improved how easy students find learning by 0.1 s.d. and reduced the self-reported anxiety while learning with friends by 0.096 s.d.. We do not find statistically significant effects on the average learning experience of students.

Based on specification 2 that disaggregates the ITT effects by tutor status of a student, we find in panels B and C of table 7 that the program effects on classroom experience are driven by effects on learners while the effects on the classroom experience of tutors follow a similar direction but

Table 6: *ITT effects on learning*

	Total	Weighted Total	Easy	Medium	Difficult
	(1)	(2)	(3)	(4)	(5)
		Panel A: All			
Treatment	0.160***	0.154***	0.061	0.088**	0.132**
	(0.054)	(0.050)	(0.038)	(0.037)	(0.055)
Control Mean	0.086	0.084	0.057	0.055	0.074
		Panel B: Learners			
Treatment	0.157***	0.152***	0.059	0.082*	0.126**
	(0.060)	(0.056)	(0.043)	(0.042)	(0.060)
Control Mean	-0.043	-0.041	-0.005	-0.025	-0.065
		Panel C: Tutors			
Treatment	0.153***	0.143***	0.063	0.089**	0.131**
	(0.053)	(0.051)	(0.040)	(0.044)	(0.059)
Control Mean	0.496	0.479	0.253	0.309	0.515
		Panel D: p-value of comparisons			
ITT: Learners vs Tutors	0.941	0.842	0.926	0.885	0.932
Observations	7966	7966	7966	7966	7966

are not statistically significant. Specifically, we find that learners report finding studying easier by 0.11 s.d. and feeling less anxious while learning with friends by 0.10 s.d.. However, they more strongly agree (by 0.09 s.d.) with the statement that poorly performing are excluded from study groups.

5.2.3 Effects on Socioemotional skills/traits

Based on our main empirical specification 1, panel A of table 8 shows that on average, the program had no statistically significant effect on any of the other measured socioemotional skills. Moreover, based on specification 2 that disaggregates the ITT effects by tutor status of a student, we find in panels B and C of table 8 that program had not effect on statistically significant on either learners' or tutor's socioemotional skills.

Figure 6: *Non-parametric Inestigation of Treatment Effects by Baseline Percentile*

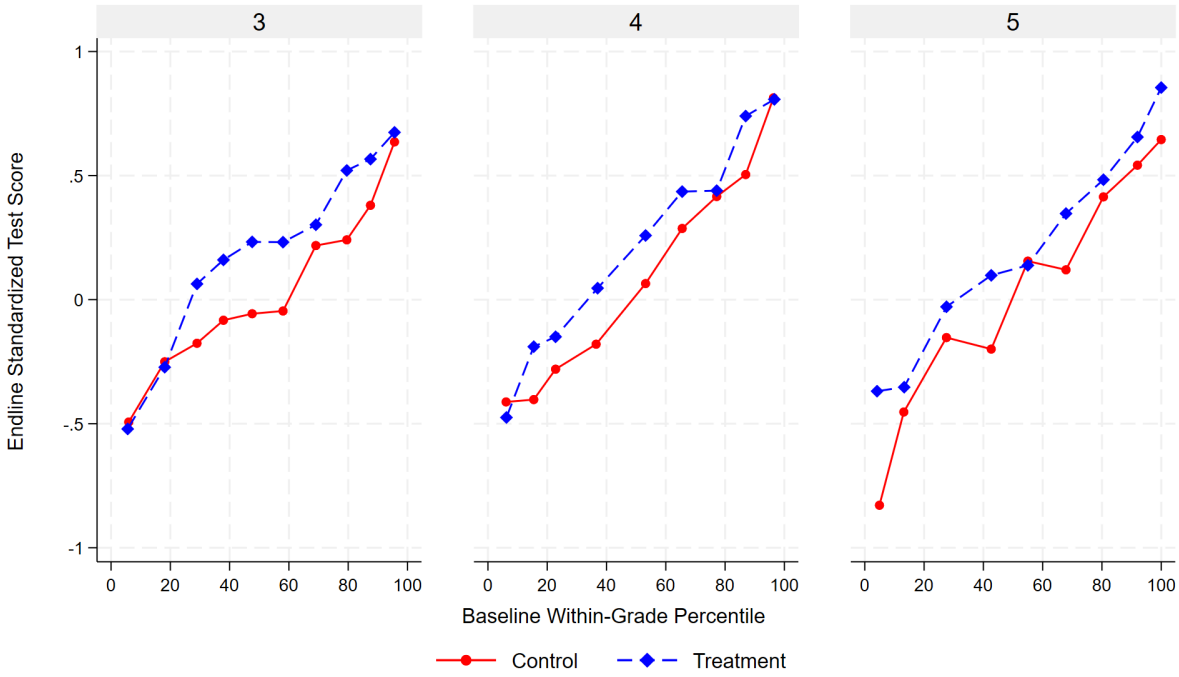


Table 7: *ITT effects on classroom experiences*

	Confident Asking Qs (1)	Studying Easy (2)	Afraid of Math (3)	Scared of HW (4)	Aspiration to Become Monitor (5)	Anxiety in Study Groups (6)	Exclude Poor Performers (7)	Befriend High Performers (8)
				Panel A: All				
Treatment	0.028	0.104**	-0.015	0.027	0.028	-0.096**	0.080	0.046
	(0.048)	(0.040)	(0.039)	(0.040)	(0.036)	(0.044)	(0.050)	(0.056)
Control Mean	-0.004	0.019	-0.003	0.009	0.051	0.012	-0.008	0.010
				Panel B: Learners				
Treatment	0.028	0.114***	0.004	0.039	0.026	-0.101**	0.085	0.054
	(0.051)	(0.041)	(0.042)	(0.043)	(0.039)	(0.049)	(0.052)	(0.056)
Control Mean	-0.046	0.002	0.030	0.045	0.018	0.063	0.006	0.009
				Panel C: Tutors				
Treatment	0.027	0.076	-0.069	-0.008	0.035	-0.088	0.062	0.023
	(0.057)	(0.060)	(0.048)	(0.055)	(0.048)	(0.055)	(0.068)	(0.080)
Control Mean	0.123	0.072	-0.101	-0.096	0.148	-0.129	-0.046	0.014
				Panel D: p-value of comparisons				
ITT: Learners vs Tutors	0.986	0.483	0.117	0.361	0.858	0.811	0.713	0.643
Observations	6686	6655	6682	6451	6730	5898	6414	6371

Table 8: *ITT effects on socioemotional skills/traits*

	Competitiveness (1)	Grit (2)	Prosociality (3)	Agreeableness (4)	Extraversion (5)	Growth Mindset (6)
			Panel A: All			
Treatment	0.034	0.020	0.050	0.012	0.016	0.020
	(0.031)	(0.045)	(0.057)	(0.051)	(0.042)	(0.033)
Control Mean	0.025	-0.004	-0.015	-0.005	0.008	0.010
			Panel B: Learners			
Treatment	0.057	0.025	0.041	-0.004	0.015	0.027
	(0.035)	(0.050)	(0.060)	(0.053)	(0.047)	(0.038)
Control Mean	-0.018	-0.050	-0.005	0.026	-0.011	-0.020
			Panel C: Tutors			
Treatment	-0.028	0.006	0.076	0.051	0.020	0.005
	(0.040)	(0.058)	(0.068)	(0.070)	(0.054)	(0.037)
Control Mean	0.154	0.130	-0.047	-0.094	0.062	0.100
			Panel D: p-value of comparisons			
ITT: Learners vs Tutors	0.051	0.739	0.543	0.410	0.931	0.613
Observations	6906	6310	6753	5759	5999	6917

5.2.4 Effects on Social Networks

Based on our main empirical specification 1, panel A of table 9 shows that on average, the program had no statistically significant effect on the size of students' network of friends or size of students' help-seeking networks. However, it reduced the number of nominations a student receives for being part of study groups by 0.18 on average (column 1) and it reduced the number of nominations a student makes as part of their study groups by 0.19 on average (column 2).

Based on specification 2 that disaggregates the ITT effects by tutor status of a student, we find in panels B and C of table 9 that learners are reported by 0.14 fewer students as part of their study group while tutors are reported by 0.27 fewer students as part of their study group (column 1). Moreover, learners report 0.20 fewer students as part of their study group while tutors report 0.15 fewer study as part of their study groups but this latter effect is not statistically significant. Taken together, this suggests that while students' number of friends or people they help or seek help from did not change, they study with fewer number of students due to the intervention.

We also study the average scores of students in these networks. Based on our main empirical specification 1, panel A of table 12 shows that on average, peers who report a student as part

Table 9: *ITT effects on social networks: size*

	In-degree Study Group (1)	Out-degree Study Group (2)	In-degree Help seeking (3)	Out-degree Help seeking (4)	In-degree Friends (5)	Out-degree Friends (6)	Named Popular (7)
			Panel A: All				
Treatment	-0.175**	-0.194***	0.035	-0.024	-0.053	-0.104	0.041
	(0.076)	(0.070)	(0.065)	(0.058)	(0.090)	(0.088)	(0.058)
Control Mean	1.532	1.399	1.473	1.320	1.784	1.632	1.227
			Panel B: Learners				
Treatment	-0.142*	-0.204***	0.023	-0.038	-0.065	-0.115	0.061
	(0.076)	(0.072)	(0.060)	(0.062)	(0.089)	(0.089)	(0.060)
Control Mean	1.410	1.343	1.287	1.277	1.690	1.577	1.093
			Panel C: Tutors				
Treatment	-0.269**	-0.154	0.087	0.027	-0.006	-0.058	-0.010
	(0.124)	(0.100)	(0.143)	(0.089)	(0.148)	(0.131)	(0.109)
Control Mean	1.918	1.578	2.062	1.456	2.082	1.806	1.653
			Panel D: p-value of comparisons				
ITT: Learners vs Tutors	0.246	0.565	0.639	0.463	0.646	0.610	0.519
Observations	7966	7966	7966	7966	7966	7966	7966

of their study group had 0.14 s.d. higher average test scores in treatment schools (column 1) while peers who a student reports as part of their study group had similar average test scores across treatment and control schools (column 2). Moreover, peers who report seeking help from a student have 0.08 s.d. higher average test scores in treatment schools (column 3) although this is not statistically significant and peers who a student reports seeking help from have 0.11 s.d. higher average test scores in treatment schools (column 4). Peers who nominate a student as their friend as well as peers who a student nominates as a friend have no statistical difference in average scores in treatment and control schools (columns 5 and 6).

Based on specification 2 that disaggregates the ITT effects by tutor status of a student, we find in panel B of table 12 that on average, peers who report a learner as part of their study group had 0.18 σ higher average test scores in treatment schools (column 1) while peers who a learner reports as part of their study group had similar average test scores across treatment and control schools (column 2). Moreover, peers who report seeking help from a learner have similar average test scores in treatment schools (column 3) and peers who a learner reports seeking help from have 0.12 s.d. higher average test scores in treatment schools (column 4). Peers who nominate a learner as their friend as well as peers who a learner nominates as a friend have no statistical difference in average scores in treatment and control schools (columns 5 and 6).

Similarly, based on specification 2 that disaggregates the ITT effects by tutor status of a student, we find in panel C of table 12 that, on average, peers who report a tutor as part of their study

Table 10: *ITT effects on social networks: average scores at endline*

	In-network Study Groups (1)	Out-network Study Groups (2)	In-network Help seeking (3)	Out-network Help seeking (4)	In-network Friends (5)	Out-network Friends (6)
			Panel A: All			
Treatment	0.143***	0.048	0.084	0.106**	0.038	0.042
	(0.050)	(0.054)	(0.056)	(0.052)	(0.053)	(0.056)
Control Mean	0.185	0.235	0.141	0.317	0.177	0.191
			Panel B: Learners			
Treatment	0.175***	0.063	0.079	0.117**	0.032	0.038
	(0.052)	(0.056)	(0.061)	(0.056)	(0.055)	(0.057)
Control Mean	0.141	0.201	0.092	0.275	0.142	0.150
			Panel C: Tutors			
Treatment	0.062	0.004	0.106*	0.073	0.063	0.059
	(0.060)	(0.068)	(0.059)	(0.060)	(0.062)	(0.065)
Control Mean	0.303	0.332	0.266	0.442	0.281	0.314
			Panel D: p-value of comparisons			
ITT: Learners vs Tutors	0.020	0.286	0.597	0.412	0.547	0.673
Observations	4588	4128	4361	4586	5003	4094

group as well as peers who a tutor reports as part of their study group had no statistically difference in average test scores in treatment and control schools (columns 1 and 2). Howeover, peers who report seeking help from a tutor have 0.11 s.d. higher average test scores in treatment schools (column 3) and peers who a tutor reports seeking help from have 0.07 s.d. higher average test scores in treatment schools (column 4) although this is not statistically significant. Peers who nominate a tutor as their friend as well as peers who a tutor nominates as a friend have no statistical difference in average scores in treatment and control schools (columns 5 and 6).

5.2.5 Heterogeneity

To better understand who benefited the most from the peer-tutoring, we used the causal forest approach developed by Athey and Imbens (2016) and Wager and Athey (2018). Following Carlana et al. (2022), we use the out-of-bag predictions to categorize students into the top and bottom half of predicted treatment effects (designated as "Weak" and "Strong" groups).

We present the conditional average treatment effect (CATE) for each subgroup (column 1), its standard error clustered at the school level (column 2), and the respective p-value (column 3). We also report the proportion of students in the "weak" group who have that characteristic (column 4) and the proportion of students in the "strong" group who have that characteristic

Table 11: *ITT effects on social networks: average scores (from baseline)*

	In-network Study Groups (1)	Out-network Study Groups (2)	In-network Help seeking (3)	Out-network Help seeking (4)	In-network Friends (5)	Out-network Friends (6)
			Panel A: All			
Treatment	-0.120*	-0.203***	-0.143*	-0.132*	-0.157**	-0.140**
	(0.063)	(0.062)	(0.077)	(0.068)	(0.079)	(0.069)
Control Mean	0.320	0.416	0.273	0.459	0.290	0.326
			Panel B: Learners			
Treatment	-0.043	-0.159**	-0.156*	-0.142*	-0.163**	-0.148**
	(0.062)	(0.065)	(0.084)	(0.075)	(0.082)	(0.073)
Control Mean	0.259	0.386	0.210	0.417	0.227	0.279
			Panel C: Tutors			
Treatment	-0.322***	-0.327***	-0.110	-0.105	-0.145	-0.117
	(0.089)	(0.078)	(0.085)	(0.080)	(0.091)	(0.080)
Control Mean	0.478	0.499	0.418	0.581	0.474	0.468
			Panel D: p-value of comparisons			
ITT: Learners vs Tutors	0.000	0.019	0.521	0.625	0.801	0.640
Observations	2820	2939	2706	3132	3255	3065

(column 5). A positive difference in column 6 indicates that the students in that subgroup are more likely to be in the "strong" group.

Focusing on student characteristics (Panels A of table 13, we find evidence that program effects are equally large for male and female students, students in higher grades benefit more and learners benefit more than tutors. Focusing on school characteristics (Panels A of table 13, we find evidence of higher effects on learning in among students in schools with large number of students and schools with fewer teachers.

Lastly, we present Targeting Operator Characteristic (TOC) curves and Rank-Weighted Average Treatment Effect (RATE) estimates, which help discern whether there is any overall heterogeneity in effects (Yadlowsky et al., 2025). The TOC curve based on the school-level clustered causal forest models presented in table 13 (Wager and Athey, 2018). The TOC curve provides the increase in the average treatment effect when treating different fractions of the sample—ranked by predicted treatment effects—compared to the overall average treatment effect if all students are treated. Dashed lines indicate 95-percent confidence intervals. The RATE, reported at the top (with standard errors in parentheses), summarizes the area under the TOC curve and serves as a formal test for the presence of heterogeneity in treatment effects (Yadlowsky et al., 2025).

Figure 7 confirms that there is meaningful heterogeneity in program effects formally. For example, if the program targeted half of the students using the student and school characteristics

Table 12: *ITT effects on social networks: average scores (network from baseline)*

	In-network Help seeking (1)	Out-network Help seeking (2)	In-network Friends (3)	Out-network Friends (4)
		Panel A: All		
Treatment	0.056	0.092	0.080	0.083
	(0.058)	(0.056)	(0.056)	(0.055)
Control Mean	0.102	0.253	0.122	0.153
		Panel B: Learners		
Treatment	0.066	0.098*	0.094	0.083
	(0.063)	(0.059)	(0.058)	(0.057)
Control Mean	0.053	0.212	0.080	0.122
		Panel C: Tutors		
Treatment	0.031	0.080	0.040	0.086
	(0.062)	(0.062)	(0.060)	(0.059)
Control Mean	0.234	0.370	0.243	0.244
		Panel D: p-value of comparisons		
ITT: Learners vs Tutors	0.522	0.701	0.193	0.947
Observations	4986	6504	6329	7069

we investigated, the average program effects would further increase by more than 0.4 standard deviations for academic learning.

6 Conclusion

We study a teacher-facilitated peer tutoring intervention in government-run primary schools in rural India one of the most resource-constrained educational environments in the world. The intervention, implemented at scale across 177 schools in partnership with district education authorities and an implementing NGO, provides compelling evidence that same-grade peer tutoring can meaningfully improve learning outcomes when embedded within existing classroom structures.

The results demonstrate five key findings. First, the intervention substantially improved mathematics performance by 0.16 s.d., with gains concentrated in medium and high-difficulty questions, indicating that peer tutoring enhanced mastery of more complex competencies rather than merely improving performance on basic skills. Second, both learners and tutors benefited equally from the program, learners improved by 0.16 s.d. and tutors by 0.15 s.d., providing

Table 13: *Exploring conditional average treatment effects on student learning*

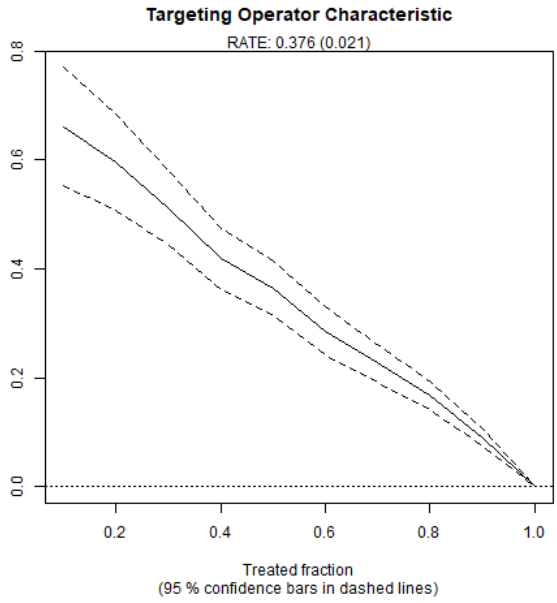
	CATE	Standard error	p-value	Weak group	Strong group	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
ITT Effect				-0.01	0.35	0.37
Panel A: Student Characteristics						
Male	0.13	0.04	0.000	44.36	40.41	-3.95
Female	0.17	0.03	0.000	54.05	58.68	4.62
Grade 3	0.07	0.04	0.098	36.03	26.25	-9.78
Grade 4	0.19	0.04	0.000	28.89	37.26	8.37
Grade 5	0.19	0.04	0.000	35.09	36.49	1.40
Leader	0.12	0.04	0.005	27.35	20.10	-7.25
Learner	0.17	0.03	0.000	72.65	79.90	7.25
Baseline test: Bottom quartile	0.11	0.06	0.047	26.31	23.67	-2.63
Baseline test: Top quartile	0.12	0.04	0.002	28.84	21.17	-7.67
Panel B: School Characteristics						
# Students: Bottom quartile	0.16	0.05	0.001	21.47	18.67	-2.81
# Students: Top quartile	0.20	0.05	0.000	27.50	36.39	8.89
# Teachers: Bottom quartile	0.13	0.04	0.003	29.83	24.47	-5.36
# Teachers: Top quartile	-0.05	0.06	0.420	33.50	14.90	-18.60
Distance to City: Bottom quartile	0.13	0.05	0.016	25.69	21.81	-3.87
Distance to City: Top quartile	0.12	0.04	0.007	26.01	21.94	-4.07

Notes. This table explores heterogeneity in treatment effects, with a focus on the average ITT effect on student learning. The sample consists of non-attriting students who sat the written exams. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

direct evidence of "learning by teaching" effects among primary school students in a real-world setting. Third, learners reported finding studying easier and experienced reduced anxiety when learning with peers, suggesting improvements in the affective dimensions of learning. Fourth, the program reshaped study networks: students formed smaller, higher-quality study groups composed of better-performing peers, indicating that peer tutoring fostered more intentional and productive academic relationships. Fifth, despite these changes in study networks, friendship networks remained unaffected, and the program did not significantly influence socioemotional skills as measured.

Heterogeneity analysis reveals that the intervention was particularly effective in schools with larger student enrollments and fewer teachersprecisely the contexts where traditional teaching methods face the greatest strain. Female students and those in higher grades benefited more, while effects were comparable across the learning distribution, indicating broad-based improvements rather than benefits accruing only to specific subgroups.

Figure 7: *Ranked Average Treatment Effect on Student Learning*



Notes. This figure explores whether there is overall heterogeneity in treatment effects along the individual student- and school-level background variables investigated in Table 13.

These findings carry important implications for education policy in low- and middle-income countries. Peer tutoring represents a cost-effective and institutionally feasible approach to improving foundational learning that does not require additional infrastructure, technology, or external personnel. By leveraging existing teachers and reorganizing classroom time, the intervention achieved effect sizes comparable to far more resource-intensive programs such as one-on-one tutoring or remedial instruction. The scalability of the approach is further demonstrated by its successful implementation in public primary schools with varying levels of teacher capacity and student populations comprising of varying learning levels. Rather than enforcing rigid group assignments, providing teachers with guidelines and allow local adaptation based on classroom knowledge was key to implementation.

Several important questions remain. First, we do not observe effects on socioemotional skills despite theoretical predictions. This may reflect measurement challenges, the short intervention period (6-7 months), or that such effects require longer time horizons to materialize. Tracking outcomes over multiple years would clarify whether academic gains translate into broader developmental benefits. Second, understanding why teachers modified groups as they did and studying whether their judgment systematically improved matching should inform program design. Analyzing the characteristics of teacher-formed groups beyond baseline ability (e.g., social dynamics, behavioral compatibility) could reveal tacit knowledge that algorithms cannot capture. Finally, testing external validity across contextsdifferent subjects, older/younger

students, urban settings, higher-capacity schoolswould establish boundary conditions for peer tutoring's effectiveness.

This study contributes to the literature on peer tutoring, active learning and scaling educational interventions in developing countries by demonstrating that peer tutoring can be effectively implemented by regular teachers in government schools without requiring wholesale pedagogical reform or teacher buy-in barriers that have hindered other interventions. It also provides rare evidence on same-grade tutoring in primary schools, extending findings from cross-age and secondary school contexts to younger learners in foundational stages of education.

In sum, this study demonstrates that peer tutoring, when thoughtfully designed and embedded within existing school systems, can serve as a powerful complement to teacher instruction in resource-constrained settings. The mutually beneficial effects for both tutors and learners, combined with the program's scalability and low cost, position peer tutoring as a promising way for addressing the learning crisis in primary education. As countries strive to meet foundational literacy and numeracy goals, interventions that reorganize existing resources rather than requiring new inputs offer a pragmatic and sustainable path forward.

References

Alan, S., Ertac, S., 2018. Fostering Patience in the Classroom: Results from Randomized Educational Intervention. Journal of Political Economy 126, 1865–1911. URL: https://www.journals.uchicago.edu/doi/full/10.1086/699007, doi:10.1086/699007. publisher: The University of Chicago Press.

ASER, 2025. Annual Status of Education Report 2024. URL: https://asercentre.org/aser-2024/.

Athey, S., Imbens, G., 2016. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences 113, 7353–7360. URL: https://www.pnas.org/doi/full/10.1073/pnas.1510489113, doi:10.1073/pnas.1510489113. publisher: Proceedings of the National Academy of Sciences.

Bank, W., 2022. Confronting the Learning Crisis: Lessons from World Bank Support for Basic Education, 201222 (Approach Paper). URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099917506172215468.

Berlinski, S., Busso, M., 2017. Challenges in educational reform: An experiment on active learning in mathematics. Economics Letters 156, 172–175. URL: https://www.sciencedirect.com/science/article/pii/S0165176517301854, doi:10.1016/j.econlet.2017.05.007.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., Wane, W., 2017. Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa. Journal of Economic Perspectives 31, 185–204. URL: https://www.aeaweb.org/articles?id=10.1257/jep.31.4.185, doi:10.1257/jep.31.4.185.

Carlana, M., La Ferrara, E., Pinotti, P., 2022. Goals and Gaps: Educational Careers of Immigrant Children. Econometrica 90, 1–29. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17458, doi:10.3982/ECTA17458. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA17458.

Chetty, R., Jackson, M.O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R.B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M., Wernerfelt, N., 2022. Social capital I: measurement and associations with economic mobility. Nature 608, 108–121. URL: https://www.nature.com/articles/s41586-022-04996-4, doi:10.1038/s41586-022-04996-4.

Deming, D.J., 2017. The Growing Importance of Social Skills in the Labor Market*. The Quarterly Journal of Economics 132, 1593–1640. URL: https://doi.org/10.1093/qje/qjx022, doi:10.1093/qje/qjx022.

Fafchamps, M., Mo, D., 2018. Peer effects in computer assisted learning: evidence from a randomized experiment. Experimental Economics 21, 355–382. URL: https://www.cambridge.org/core/journals/experimental-economics/article/abs/peer-effects-in-computer-assisted-learning-evidence-from-a-randomized-experiment/A941DD955FA4F93A9DC7E854AA54A0F5, doi:10.1007/s10683-017-9538-z.

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., Wenderoth, M.P., 2014. Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences 111, 8410–8415. URL: https://www.pnas.org/doi/10.1073/pnas.1319030111, doi:10.1073/pnas.1319030111. publisher: Proceedings of the National Academy of Sciences.

Fryer, R.G., 2017. Chapter 2 - The Production of Human Capital in Developed Countries: Evidence From 196 Randomized Field Experimentsa, in: Banerjee, A.V., Duflo, E. (Eds.), Handbook of Economic Field Experiments. North-Holland. volume 2 of *Handbook of Economic Field Experiments*, pp. 95–322. URL: https://www.sciencedirect.com/science/article/pii/S2214658X16300083, doi:10.1016/bs.hefe.2016.08.006.

Gal, A., Fallik, O., 2021. Learn from Each Other: A Peer-Teaching Model. Interdisciplinary Journal of Environmental and Science Education 17, e2242. URL: https://www.ijese.com/article/learn-from-each-other-a-peer-teaching-model-10896, doi:10.21601/ijese/10896.

Glewwe, P., Muralidharan, K., 2016. Improving Education Outcomes in Developing Countries, in: Handbook of the Economics of Education. Elsevier. volume 5, pp. 653–743. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780444634597000105, doi:10.1016/B978-0-444-63459-7.00010-5.

GoI, 2010. Right to Education. URL: https://dsel.education.gov.in/en/rte.

GoI, 2020. National Education Policy 2020.

Kraft, M.A., Schueler, B.E., Falken, G., 2024. What Impacts Should We Expect from Tutoring at Scale? Exploring Meta-Analytic Generalizability. EdWorkingPaper No. 24-1031. Technical Report. Annenberg Institute for School Reform at Brown University. URL: https://eric.ed.gov/?id=ED661550. eRIC Number: ED661550.

Li, T., Han, L., Zhang, L., Rozelle, S., 2014. Encouraging classroom peer interactions: Evidence from Chinese migrant schools. Journal of Public Economics 111, 29–45. URL: https://www.sciencedirect.com/science/article/pii/S0047272713002569, doi:10.1016/j.jpubeco.2013.12.014.

List, J., 2022. The Voltage Effect by John A. List: 9780593239483 | PenguinRandomHouse.com: Books. URL: https://www.penguinrandomhouse.com/books/672117/the-voltage-effect-by-john-a-list/.

Nickow, A., Oreopoulos, P., Quan, V., 2020. The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. Technical Report w27476. National Bureau of Economic Research. URL: https://www.nber.org/papers/w27476, doi:10.3386/w27476.

Romero, M., Chen, L., Magari, N., 2022. Cross-Age Tutoring: Experimental Evidence from Kenya. Economic Development and Cultural Change 70, 1133–1157. URL: https://www.journals.uchicago.edu/doi/10.1086/713940, doi:10.1086/713940. publisher: The University of Chicago Press.

Shah, R., 2025. When the Student Becomes the Master: A Field Experiment on Learning by Teaching. Job Market Paper URL: https://www.dropbox.com/scl/fi/iabm2hw5f5980osxgikee/7-30-25_Shah.pdf?dl=0&rlkey=duseln36g9y43v4zy09p27ul5.

Sinha, S., Banerji, R., Wadhwa, W., 2016. Teacher Performance in Bihar, India: Implications for Education. doi:10.1596/978-1-4648-0739-8.

Wager, S., Athey, S., 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. Journal of the American Statistical Association 113, 1228–1242. URL: https://doi.org/10.1080/01621459.2017.1319839, doi:10.1080/01621459.2017.1319839.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., Wager, S., 2025. Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects. Journal of the American Statistical Association 120, 38–51. URL: https://doi.org/10.1080/01621459.2024.2393466, doi:10.1080/01621459.2024.2393466. publisher: ASA Website _eprint: https://doi.org/10.1080/01621459.2024.2393466.

Appendix

A Additional figures and tables

Figure A1: *Peer Group Visuals Inside Classrooms*

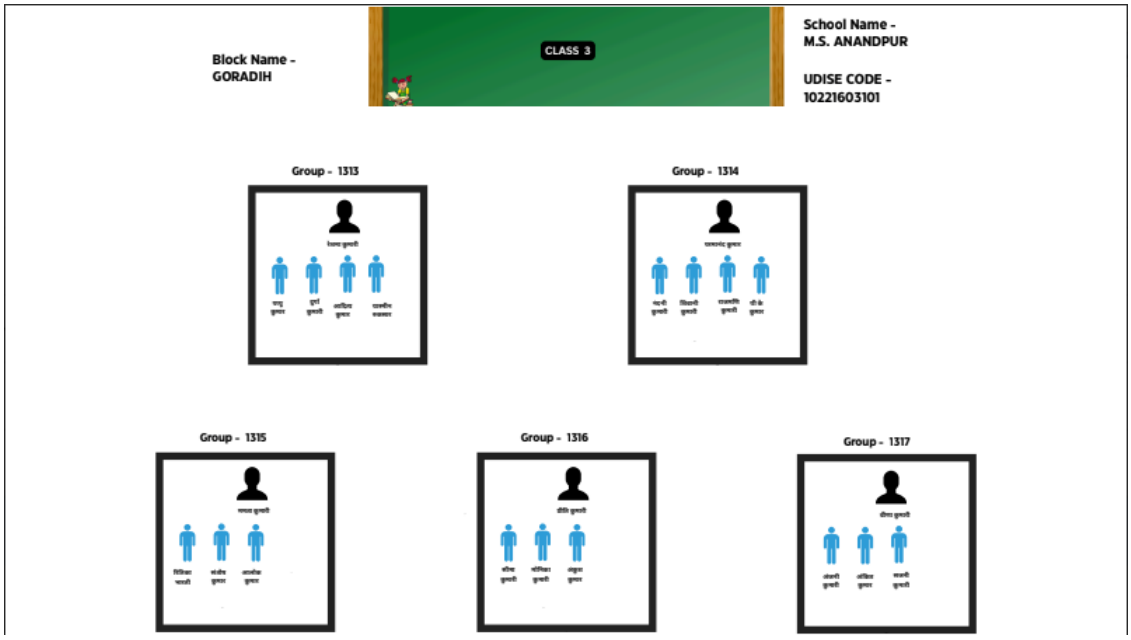


Figure A2: *Peer Groups Inside Classrooms*



Figure A3: *Timeline*

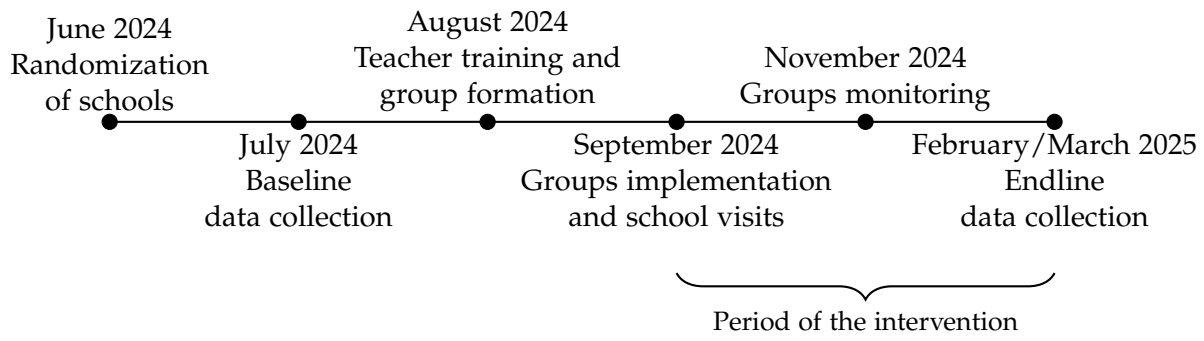


Figure A4: *Cumulative distribution of learning levels at baseline by grade*

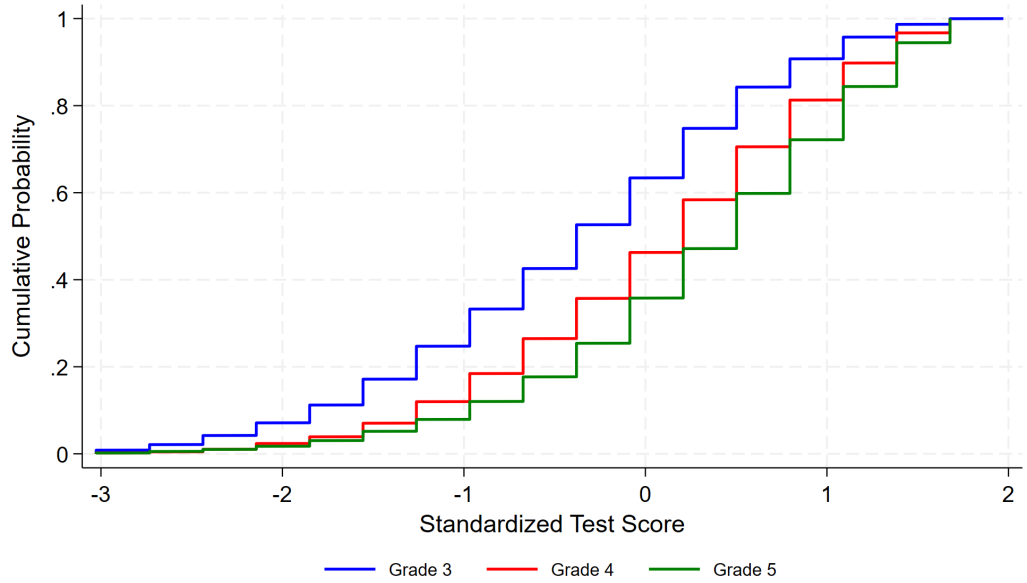


Figure A5: *Distribution of fraction of students in each grade in a school in a decile*

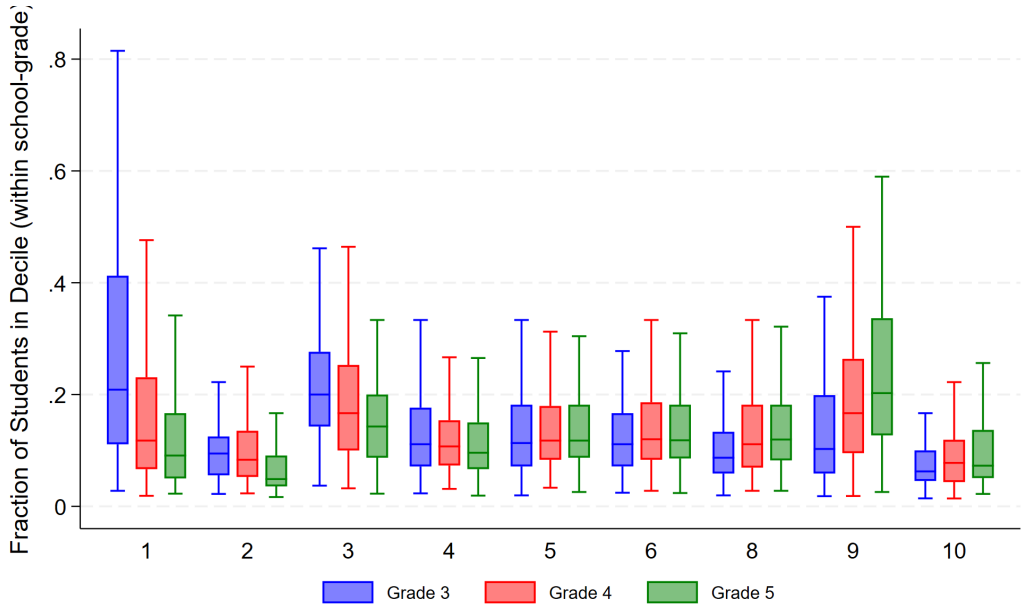


Figure A6: *Grade-wise distribution of average scores of "incoming links" for friendship networks*

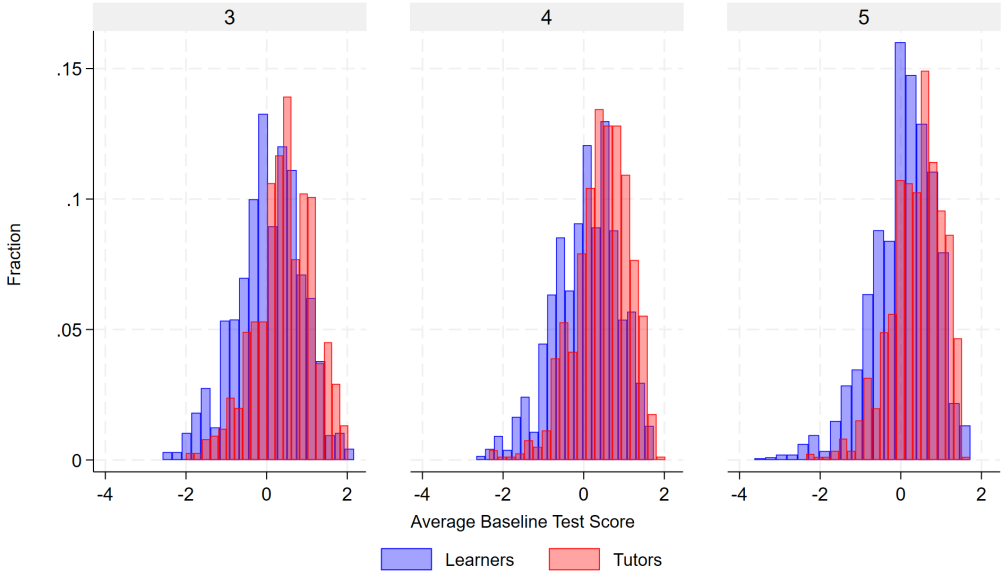


Figure A7: *Grade-wise distribution of average scores of "incoming links" for help-seeking networks*

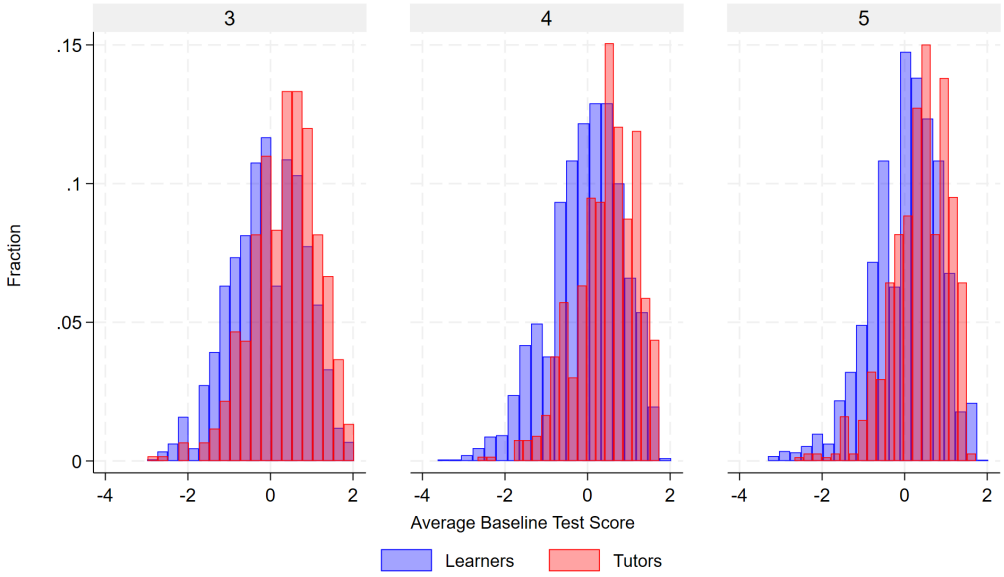


Figure A8: *Grade-wise distribution of average scores of "incoming links" for friendship networks*

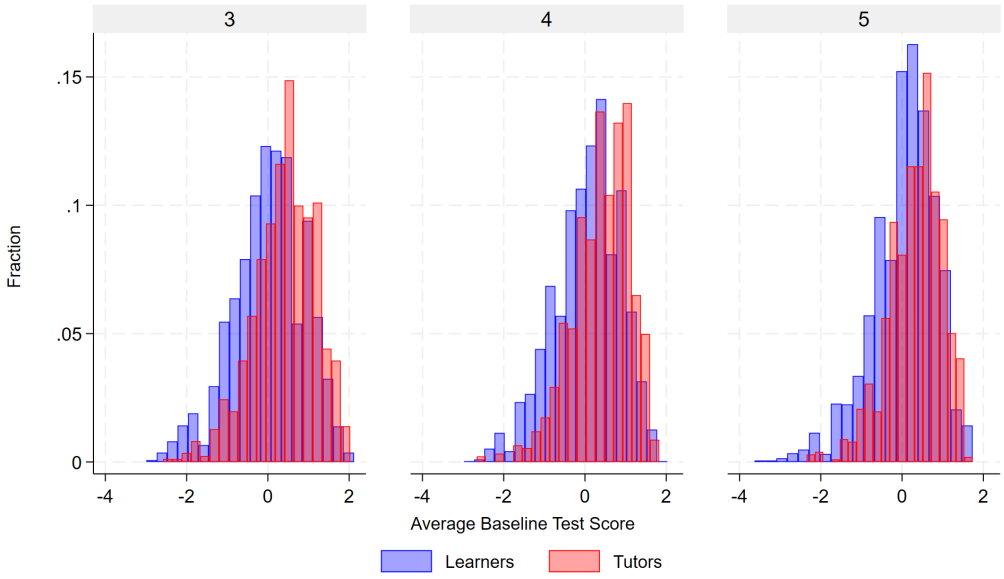


Figure A9: *Grade-wise distribution of average scores of "incoming links" for help-seeking networks*

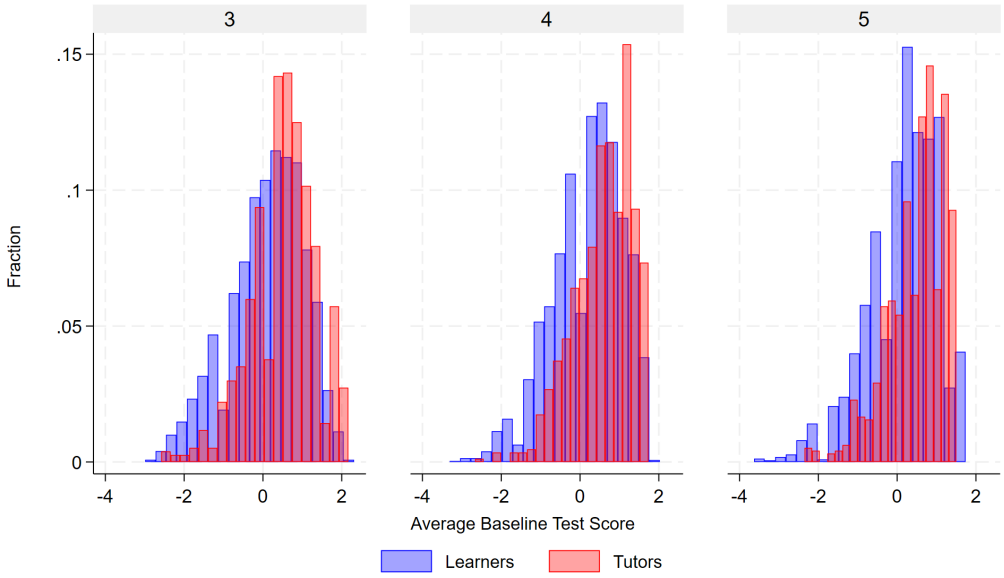
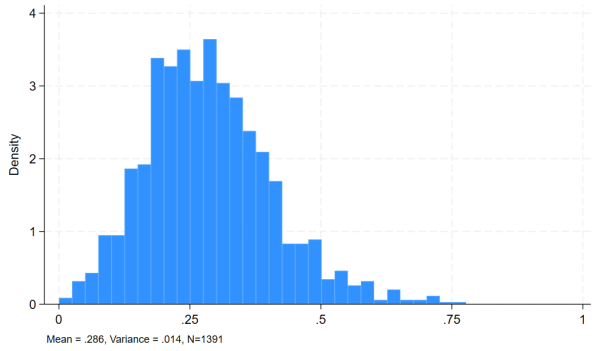


Figure A10: *Distribution of Average Distance in Math Scores Between Tutors and Learners in a Group*

(a) *Assigned Groups*



(b) *Implemented Groups*

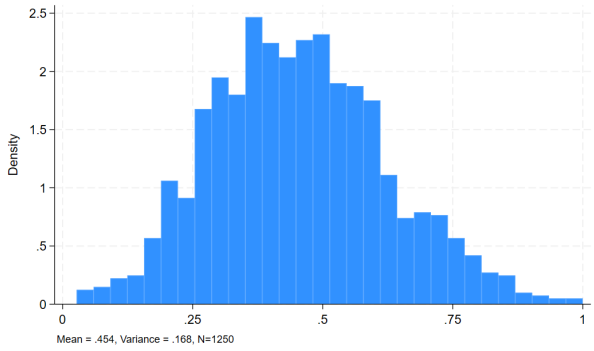
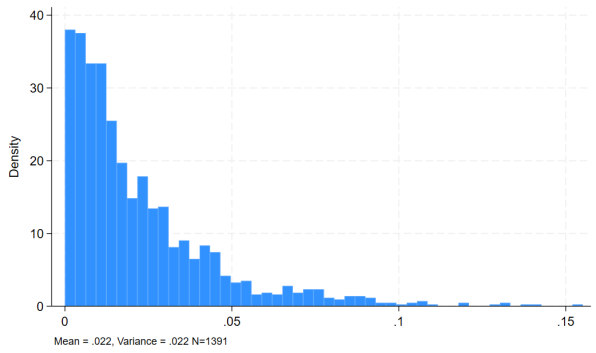


Figure A11: *Distribution of Standard Deviation of Distance in Math Scores Between Tutors and Learners in Group*

(a) *Assigned Groups*



(b) *Implemented Groups*

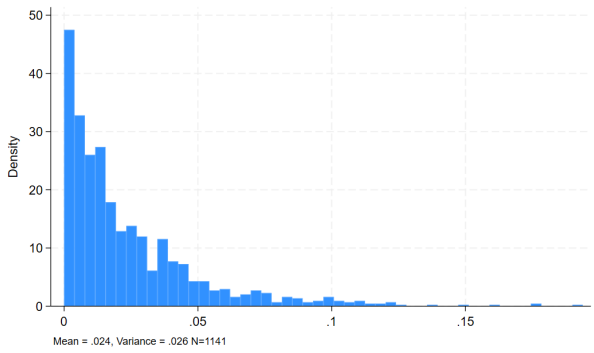
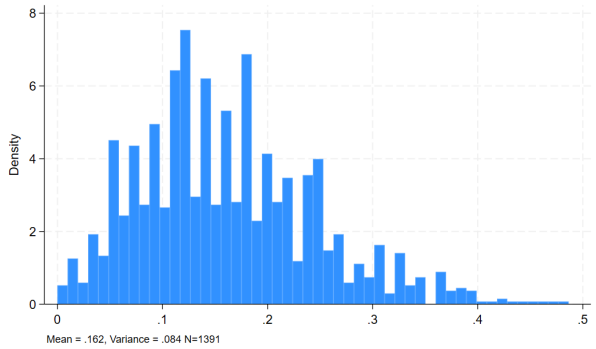


Figure A12: *Distribution of Average Distance in Math Scores Between Learners in Group*

(a) *Assigned Groups*



(b) *Implemented Groups*

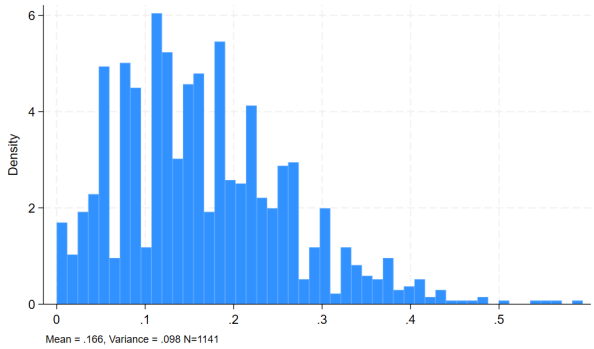
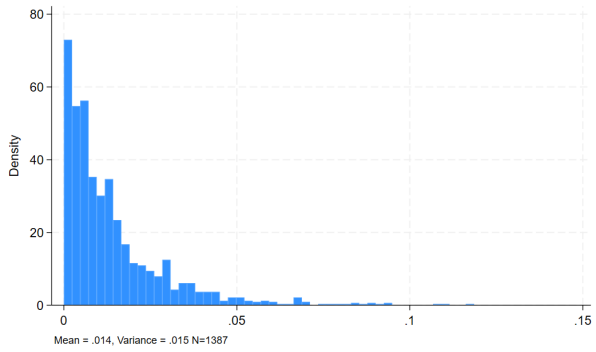


Figure A13: *Distribution of Standard Deviation of Distance in Math Scores Between Learners in Group*

(a) *Assigned Groups*



(b) *Implemented Groups*

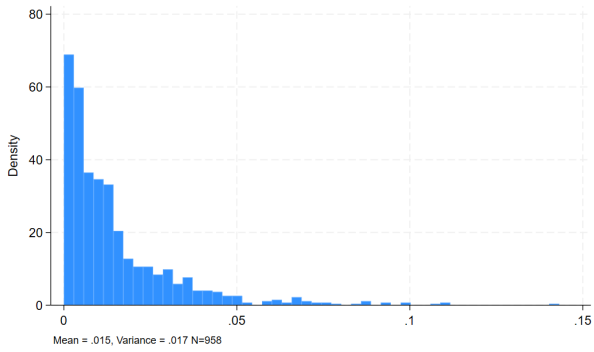


Table A1: *Descriptive statistics and balancing checks: I*

	Number of observations		Balancing check		
	Control	Treatment	Control	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)
Panel A: Student Characteristics					
Age	7473	6349	9.65	9.64	0.00
			[1.27]	[1.23]	(0.04)
Female	7425	6296	0.56	0.55	-0.01
			[0.50]	[0.50]	(0.01)
Attrited	7539	6396	0.42	0.42	-0.00
			[0.49]	[0.49]	(0.02)
Joint F-test (p-value)					0.91
Panel B: Student Math Skills					
Addition (Standardized)	7539	6396	0.02	-0.03	-0.06
			[1.00]	[1.03]	(0.04)
Subtraction (Standardized)	7539	6396	0.03	-0.03	-0.09*
			[1.00]	[1.01]	(0.05)
Multiplication (Standardized)	7539	6396	0.01	-0.02	-0.05
			[1.00]	[1.02]	(0.05)
Division (Standardized)	7539	6396	0.01	-0.02	-0.04
			[1.00]	[0.98]	(0.05)
Pictorial (Standardized)	7539	6396	0.05	-0.06	-0.14**
			[1.00]	[1.03]	(0.06)
Counting (Standardized)	7539	6396	0.07	-0.09	-0.18***
			[1.00]	[1.02]	(0.06)
Word Problems (Standardized)	7539	6396	0.06	-0.07	-0.15**
			[1.00]	[1.07]	(0.07)
Joint F-test (p-value)					0.40

Notes. This table describes the study's sample of 177 schools and presents balance checks. Control refers to the 95 schools not assigned to receive the peer tutoring intervention, and Treatment refers to 82 schools assigned to receive this intervention. Balancing check reports on the regression-adjusted difference. Standard deviations are shown in brackets; standard errors are shown in parentheses. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: *Descriptive statistics and balancing checks: II*

	Number of observations		Balancing check		
	Control	Treatment	Control	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)
Panel A: Student Socioemotional Wellbeing					
Shy in asking questions	6868	5845	-0.06	0.07	0.15***
			[1.00]	[1.04]	(0.05)
Find studying easy	6863	5846	0.01	-0.01	-0.05
			[1.00]	[1.00]	(0.05)
Afraid of Math	6893	5880	-0.05	0.06	0.12***
			[1.00]	[1.03]	(0.05)
Do not enjoy homework	6721	5809	-0.04	0.06	0.12***
			[1.00]	[1.03]	(0.05)
Curious	6689	5790	0.00	-0.01	-0.01
			[1.00]	[0.99]	(0.04)
Important to score good grades	6991	5965	0.01	-0.02	-0.04
			[1.00]	[1.01]	(0.04)
Hardwork will improve grades	6999	5963	0.03	-0.03	-0.06
			[1.00]	[1.02]	(0.05)
Want to become monitor	6942	5896	0.03	-0.04	-0.08*
			[1.00]	[1.04]	(0.04)
Joint F-test (p-value)					0.12
Panel B: Student Social Networks					
In-degree (Help in Studies)	7539	6396	1.49	1.49	0.00
			[2.19]	[2.20]	(0.07)
Out-degree (Help in Studies)	7539	6396	1.49	1.49	-0.00
			[1.27]	[1.33]	(0.07)
In-degree (Friends)	7539	6396	2.40	2.46	0.06
			[2.52]	[2.71]	(0.11)
Out-degree (Friends)	7539	6396	2.40	2.46	0.06
			[1.91]	[2.07]	(0.11)
Named Popular	7539	6396	1.20	1.28	0.08
			[1.89]	[1.97]	(0.06)
Joint F-test (p-value)					0.12

Notes. This table describes the study's sample of 177 schools and presents balance checks. Control refers to the 95 schools not assigned to receive the peer tutoring intervention, and Treatment refers to 82 schools assigned to receive this intervention. Balancing check reports on the regression-adjusted difference. Standard deviations are shown in brackets; standard errors are shown in parentheses. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: *ITT effects on learning: II*

	Addition (1)	Subtraction (2)	Multiplication (3)	Division (4)	Picture problems (5)	Numbers (6)	Word Problems (7)
			Panel A: All				
Treatment	0.034	0.026	0.065	0.076**	0.052	0.103	0.138*
	(0.023)	(0.037)	(0.043)	(0.037)	(0.038)	(0.081)	(0.079)
Control Mean	0.055	0.049	0.057	0.047	0.035	0.046	0.075
			Panel B: Learners				
Treatment	0.030	0.018	0.062	0.078*	0.047	0.081	0.151*
	(0.028)	(0.041)	(0.047)	(0.041)	(0.043)	(0.087)	(0.086)
Control Mean	0.010	-0.027	-0.034	-0.049	-0.017	-0.057	-0.019
			Panel C: Tutors				
Treatment	0.048	0.049	0.074	0.068	0.059	0.139	0.081
	(0.030)	(0.043)	(0.053)	(0.042)	(0.042)	(0.085)	(0.082)
Control Mean	0.197	0.289	0.344	0.352	0.199	0.374	0.373
			Panel D: p-value of comparisons				
ITT: Learners vs Tutors	0.633	0.491	0.813	0.816	0.797	0.452	0.342
Observations	7966	7966	7966	7966	7966	7966	7966