

Asim, Minahil; Jain, Ronak; Khandelwal, Vatsal

Working Paper

Great expectations? Experimental evidence from schools in Pakistan

Working Paper, No. 454

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Asim, Minahil; Jain, Ronak; Khandelwal, Vatsal (2024) : Great expectations? Experimental evidence from schools in Pakistan, Working Paper, No. 454, University of Zurich, Department of Economics, Zurich,
<https://doi.org/10.5167/uzh-262620>

This Version is available at:

<https://hdl.handle.net/10419/306337>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich**^{UZH}

University of Zurich
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 454

Great Expectations? Experimental Evidence from Schools in Pakistan

Minahil Asim, Ronak Jain and Vatsal Khandelwal

September 2024

Great Expectations?

Experimental Evidence from Schools in Pakistan

Minahil Asim *

Ronak Jain [†]

Vatsal Khandelwal [‡]

September 30, 2024

Abstract

We study the effect of communicating student-specific teacher expectations on academic performance. We randomize whether students (a) receive high-performance expectations, (b) are additionally paired with a classmate for encouragement, (c) receive information about past performance, or (d) receive no message. Expectations increase math scores by 0.19σ , with especially large effects among students who randomly received ambitious expectations and were predicted to perform poorly. Information provision has comparably large effects (0.16σ), particularly in schools with low parental literacy. However, pairing students only improves scores when peers have similar characteristics. Our findings highlight low-cost, sustainable ways of leveraging teachers to improve performance.

Keywords: Expectations, Information, Peer Effects, Motivation, Performance.

JEL code(s):D91, D84, I24, I25, C93, D83.

*University of Ottawa. Email: minahil.asim@uottawa.ca

[†]University of Zurich. Email: ronak.jain@econ.uzh.ch

[‡]University of Exeter. Email: v.khandelwal@exeter.ac.uk

⁰We are grateful to our partner schools for their collaboration and facilitation of this research, especially Scherezade Tarar and Farah Rehman. This research has been possible because of financial support from RISE Pakistan and JPAL-PPE. We thank the research team at the Center for Economic Research in Pakistan (CERP), especially Um-mara Waheed and Waheed Ahmad for field management, and Absar Ali and Omar Zahid for exceptional research assistance. We are also grateful to Tahir Andrabi, Emily Breza, Christina Brown, Jishnu Das, Stefan Dercon, Asim Khwaja, Simon Quinn, and Gautam Rao for their useful comments and suggestions. We also thank seminar participants at Pacific Conference for Development Economics (PacDev), Association for Education Finance and Policy (AEFP) Annual Conference, LEAPS Pakistan Conference, the AEFP EdDev Community Group, Lahore University of Management Sciences, Harvard University and the University of Oxford. IRB approval was obtained from Harvard University and the University of Oxford. Our partner school chain's internal research ethics board also reviewed and approved our research protocols and consent forms in November 2020 in addition to our university Institutional Review Boards. The RCT was pre-registered as AEARCTR-0007846 at <https://www.socialscisceregistry.org/trials/7846>.

1 Introduction

We currently face a global learning crisis, with 250 million children lacking basic literacy and numeracy skills despite many of them being in school (Crouch et al., 2021; UNESCO, 2014). Existing evidence shows that “effective” teachers can improve student learning (Bau and Das, 2020; Beteille and Evans, 2019; Chetty et al., 2014a,b; Evans and Popova, 2016; World-Bank, 2017), and emphasizes the need for student-centered teaching practices where every learner feels valued (UNESCO, 2020). Communication of personalized high-performance expectations from teachers to *all* students in the classroom can be one such inclusive practice that potentially improves student outcomes.¹ Research in education psychology also hypothesizes a self-fulfilling prophecy between teacher expectations and student performance, known as the “Pygmalion Effect” (Rosenthal and Jacobson, 1968; Wang et al., 2018). Typically, however, teachers do not explicitly communicate their expectations to students, or selectively choose to convey them to those already performing well (Wolf and Brown, 2023). Moreover, it is not ex-ante clear whether high expectations can have a meaningful effect on student achievement as students could be motivated or frustrated by them (Friedrich et al., 2015), may not have a supportive classroom peer environment (Bursztyn et al., 2019), and there might be little perceived added value beyond receiving information about one’s performance (Andrabi et al., 2017; Barrera-Osorio et al., 2020).

In this paper, we experimentally investigate whether personalized high expectations of academic performance and effort, conveyed by teachers, can improve student achievement. We incorporate variations in our experiment design to evaluate how this effect depends on the magnitude of expectations, whether it can be strengthened by peer encouragement, and how it compares to simply receiving information about previous performance. We partnered with a large private school chain in Pakistan for this study. Teachers are held in high regard in our setting, as in many other contexts (Dozza and Cavrini, 2012; Wentzel, 2010). For example, 84% of parents in our sample report that it would be ‘very’ or ‘extremely’ useful if the teacher set and conveyed a concrete expectation of performance to their child. Despite this, only 23% of the teachers prefer to communicate expectations to low-performing students as opposed to those who are already performing well.

Working with 288 classrooms across 15 schools, we first collected realistic, student-specific teacher expectations. To do so in a standardized manner, we used administrative data to remind math teachers of the most recent math test scores of each of their students.² Then, all teachers were asked to fill in responses to the following statements: (1) “I expect the student to

¹In this context, the word ‘expectation’ is not used in the statistical sense of prediction but rather in its commonly used form to indicate desired outcomes.

²This helped mitigate concerns about biases in teacher expectations of student performance (Djaker et al., 2022) and we do not find any significant differences between teacher expectations across student gender, parental wealth, or age (Figure A.4.1).

work hard and improve to *achieve at least X (out of 100%)* in upcoming exams and tests" (which we refer to as 'High Expectations' hereafter) and (2) "I expect the student to work hard and improve and I think that *even Y (out of 100%) is achievable* in upcoming exams and tests" (which we refer to as 'Very High Expectations' hereafter).^{3 4}

Following this, we divided the sample classrooms randomly into three groups: an Expectations Arm, a Peer Arm, and a Comparison Group which was further split into an Information Arm and a Pure Control Arm. Students in the Expectations Arm randomly received one of the two statements of teacher expectations ('High Expectations' or 'Very High Expectations'), along with a custom-designed and individually tailored infographic encouraging them to work towards achieving the expectation.⁵ We designed the infographic as a visual aid to reinforce high expectations and encourage student engagement with the message. Students in the Peer Arm received the same message as the Expectations Arm but were also randomly matched with another classmate and received an additional infographic highlighting that they should encourage each other to achieve their respective goals. Students in the Information Arm received an infographic containing information on their performance in the previous math exam. Finally, students in the Control Arm received no message. We used the school's Google Classroom infrastructure to send private emails to each student. All teachers were blind to the treatment status of classrooms and students to ensure that they did not selectively change their efforts towards any students.

Our experiment design addresses various challenges in causally measuring the impact of high teacher expectations on student outcomes. It also allows us to uncover several important, policy-relevant insights. First, the comparison of the Expectations Arm to the Control Group allows us to evaluate the effect of communicating high teacher expectations. Moreover, random variation in the type of high teacher expectation statement, i.e. giving the expectation as a lower bound ("...to achieve at least X") or an ambitious score ("...even Y is achievable") allows us to investigate whether the magnitude of expectation and the way of conveying it influences how effective it is. Second, the Information Arm allows us to measure the effect of simply reminding students about their performance and gives us a relevant benchmark to compare the treatment effect of expectations. Third, the Peer Arm allows us to measure the effect of randomly matching classmates to encourage each other in addition to receiving teacher expectations. Random matching within the Peer Arm additionally allows us to measure the causal effect of being matched to a peer with a specific set of characteristics.

³This information was collected from all teachers before the assignment of classrooms to control or treatment status. All teachers were informed that these expectations might be communicated to students.

⁴It is important to note that the statements were designed to elicit a lower bound and an ambitious score that the teacher wanted the student to work hard toward and achieve.

⁵Figure 2 shows the infographics. The infographic displayed their previous score as the starting point and the score in the teacher's expectation as a goal to work towards.

Our experiment generates three main findings.⁶ First, we find that students in the Expectations Arm scored 0.19σ higher in their math exam than students in the Control Group who received no information. The effect is statistically significant at 1%.⁷ We find that the treatment effect is large (0.23σ) and statistically significant at 1% for students who randomly received the ‘Very High’ teacher expectation statement.⁸ We also find evidence that the magnitude of the expectation matters. In particular, we find that a 10 percentage point increase in the gap between teacher expectations and student’s baseline score leads to a 0.1σ increase in the impact of the Expectations Arm (statistically significant at 1%). Together, these findings suggest that students respond especially well to ambitious expectations communicated by the teacher. This is also confirmed by a follow-up survey where the majority of students reveal that they interpret the expectation as their teacher setting a goal to encourage them. We hence show that providing students with ambitious goals set by teachers does not frustrate them, but rather has a positive effect on their academic outcomes. Moreover, we also find that those predicted to perform the worst, score 0.33σ higher in the Expectations Arm compared to comparable students in the Control Group. This shows that the effect is not concentrated among those who would perform well in the absence of the treatment. Further, we also do not detect any negative spillover effects on English test scores.

Second, we find that students in the Information Arm also score 0.16σ higher than students in the Control Group and this effect is not significantly different from the effect of the Expectations Arm.⁹ Similar to the Expectations Arm, we continue to find that the treatment effect of the Information Arm is 0.30σ among students who are predicted to perform the worst. However, unlike the Expectations Arm, we find that the treatment effect of the Information Arm is significantly higher in schools with low parental literacy compared to schools with high parental literacy. This suggests that the simple infographic might have made the information on student performance more accessible to and easily understood by parents in settings with low literacy. Our follow-up survey also suggests that students found the image accessible, the score easier to remember, and interpreted the reminder as an encouragement from the teacher. We rule out other factors, such as delivering the performance information in a different format (i.e. as a percentage), or the timing of the delivery as alternative mechanisms.

⁶In Section B.2, we also present a simple effort-choice model to interpret how expectations, information, and peer comparisons can affect extrinsic motivation and achievement.

⁷These estimations pool results from two rounds of high-stakes math exams, at the end of the academic year (two to four weeks after the delivery of teacher expectations) and mid-year of the next academic year (six months after the delivery of teacher expectations). We also measure the treatment effects separately and find that these are 0.20σ (statistically significant at 5%) and 0.17σ at (statistically significant at 10%) respectively.

⁸The effect of the ‘High’ Expectation statement was 0.12σ . The p-value for the comparison of the treatment effects of the ‘High’ and ‘Very High’ statements in the Expectations Arm is 0.12.

⁹We also find similar treatment effects of the Information Arm at midline (0.21σ) and endline (0.11σ). While the effects are smaller and insignificant in the endline, the difference in treatment effects at midline and endline is not statistically significant. Moreover, the treatment effect of the Expectations Arm and Information Arm is not statistically distinguishable at both midline and endline.

Third, we find that on average, the treatment effect of the Peer Arm is not statistically distinguishable from zero. Even though the Peer Arm adds a peer-matching component to the Expectations Arm, we find that its effect is significantly lower than the effect of expectations alone. However, we employ the method in [Chernozhukov et al. \(2018\)](#) and find evidence of substantial heterogeneity. In particular, the effect is positive and significantly higher for matched peers who are similar to each other in terms of their baseline characteristics including their scores and teacher expectations. By contrast, it is negative and significantly lower for those whose matched peer has a higher baseline score or expectation than them. Our follow-up survey also shows that students are likely to feel disappointed if their matched peer's teacher expectation is higher than their own. On the contrary, the majority reported that they would feel motivated and happy if they were paired with a peer with similar scores and teacher expectations. Overall, the results suggest that students may feel encouraged if paired with similar peers, and discouraged if paired with dissimilar peers. These relative comparisons and morale effects likely explain the insignificant average treatment effect.

Our paper makes three key contributions. First, our study experimentally evaluates the impact of communicating high teacher expectations and benchmarks it against the impact of information. While the idea that a teacher's expectations can affect students performance leading to a self-fulfilling prophecy is not new ([Friedrich et al., 2015](#); [Rosenthal and Jacobson, 1968, 1992](#); [Wang et al., 2018](#)), there have been few experimental studies which test and quantify the extent to which this affects student performance. This is because expectations are endogenous and selectively delivered, making it difficult to evaluate their causal impact ([Carlana, 2019](#); [Jussim and Harber, 2005](#)). Additionally, while high expectations have been cited to play an important role in other contexts, for example as one of the bundled components of the US Charter Schools system ([Angrist et al., 2013](#); [Fryer Jr, 2014](#)), their role as a stand-alone component has not been studied in existing literature, especially in developing country contexts.

Additionally, our study complements the literature on the provision of information about student performance to parents ([Andrabi et al., 2017](#); [Barrera-Osorio et al., 2020](#)). We show that providing information about past performance to students can have a positive effect as large as the effect of conveying expectations. This is especially the case in settings with lower parental literacy. However, in settings where information provision or interpretation is not a binding concern, setting and communicating high teacher expectations can indeed be an effective strategy to motivate students in a sustained way. Moreover, we show that both information and expectations are especially helpful for low-performing students. Our design therefore allows us to compare the effects of these two distinct types of communication — expectations and information — between teachers and students.

Second, we analyze the effects of a low-cost, non-invasive, and sustainable intervention that can be easily implemented in low-resource settings. Prior interventions have focused on a

variety of approaches ranging from increasing student and teacher incentives (Berry, 2015; De Ree et al., 2018; Hirshleifer, 2015; Kremer et al., 2009; Mbiti et al., 2019) to reducing non-cognitive constraints by promoting the growth mindset or improving grit (Alan et al., 2019; Duckworth and Quinn, 2009; Ganimian, 2020; Islam et al., 2021; Yeager and Dweck, 2012). The evidence on the effectiveness of the interventions is mixed and context-dependent. Moreover, these interventions can be difficult for schools to naturally sustain as they are both resource and time-intensive. By contrast, we find that our intervention is about 10 times more cost-effective compared to several other recent interventions that have been effective in raising student performance in developing countries (Beteille and Evans, 2019; Glewwe and Muralidharan, 2016).¹⁰ Some studies explore other low-cost approaches to improve student performance such as allowing students to set their own goals and giving frequent nudges about working towards them (Damgaard and Nielsen, 2018; Dobronyi et al., 2019; Morisano et al., 2010; Oreopoulos and Petronijevic, 2019; Schippers et al., 2015). However, these studies do not leverage the motivational power of teacher expectations to influence student goals and their drive to achieve them.

Finally, we contribute to the literature on peer effects in education by studying how pairing students to motivate each other can affect the impact of teacher expectations. There is a growing body of research highlighting the impact of classroom norms (Bursztyn et al., 2019; Bursztyn and Jensen, 2015) and peer effects (Burke and Sass, 2013; Calvó-Armengol et al., 2009; Wu et al., 2023) on student effort and performance. We contribute to this literature by causally identifying peer effects and showing how student outcomes depend on the characteristics of their randomly matched peer, including their teacher expectations and baseline performance. Our finding that being matched with similar peers can strengthen the impact of teacher expectations is also relevant to research on the impact of classroom homophily, i.e. similarity of student characteristics, on student outcomes. For example, Jackson et al. (2023) find that only female students benefit from being matched with academically stronger peers. Our results on the effect of peer similarity yield important implications on teaching practices ranging from the formation of study groups within the class to using peers to improve student motivation and performance.

We proceed as follows. In section 2, we describe the empirical setting. We present the experiment design in Section 3 and describe the empirical strategy in Section 4. Section 5 presents the results and Section 6 provides a detailed discussion of the mechanisms. Section 7 concludes.

¹⁰We document the details of our cost-benefit analysis in Section 6.

2 Data

2.1 Setting

The education system in Pakistan includes public, low-cost private, and private schools. The incidence of private schools has grown rapidly over the years with one out of every three children found to be studying in a private school ([Andrabi et al., 2007](#)). We partnered with a private school chain that operates approximately 300 schools across Pakistan, catering to middle- and upper-middle-income families. The schools have pre-primary (KG), primary (grades 1-5), lower-secondary (grades 6-8), and secondary (grades 9-11) levels of education.

We conducted our study in a geographically spread out sample of 15 schools with grades 3 to 8 across 288 classrooms. Appendix Figure [A.1.1](#) plots the geographic location of schools in our sample. Approximately 44% of the schools cater to upper-middle income groups, while 38% to middle-income groups. About 13, 75, and 13% of the schools have low, medium, and high levels of parental literacy, respectively. 13% of schools have parents who are not comfortable with technology while 88% of schools have parents who are only somewhat comfortable with technology. This suggests that there is considerable variation in student backgrounds within our sample. Overall, our sample constitutes 1,537 students across 288 classrooms and 118 Mathematics teachers teach in these classes.

2.2 Data Sources

2.2.1 Academic Achievement Data

Each academic year has two terms, August to December and January to June. High-stakes standardized tests in Math and English are administered in every grade once every term. We collected administrative data from our partner schools, which included test scores for Math and English from three points in time: (1) December 2020 (used for stratified randomization), (2) June 2021 (end-term test scores following our intervention), and (3) December 2021 (at the end of the second term during our study). These standardized tests are designed by our partner schools' head office, reflect the curriculum being taught in different grades, and are high stakes as they determine progression onto the next grade.

2.2.2 Student and Teacher Surveys

We conducted three rounds of online surveys with students: baseline in March 2021, midline in June 2021, and endline in September-October 2021. These surveys measured classroom effort, peer networks, classroom behavior, attitudes, and non-cognitive outcomes such as motivation,

grit, and growth mindset. Out of 1537 students in our sample, 721 and 752 students (i.e., less than 50%) responded to the surveys at midline and endline respectively.¹¹ In addition to these three rounds of surveys, we conducted a follow-up student survey a year after the intervention ended to gather additional information on how students interpreted various components of the information provided to them. 997 students (65%) responded to this survey.

We also conducted two rounds of surveys with the teachers at baseline (February 2021) and endline (November 2021). These surveys measure teacher beliefs about student performance, and classroom-level outcomes such as engagement, disruptions, and warnings teachers gave to students. 118 teachers responded to the baseline survey, while only 67 teachers attempted the survey at endline.¹²

2.2.3 Data on Teacher Expectations

We elicited expectations from teachers about each of their students' math performance. To collect these in a standardized way, we asked teachers to share **realistic** expectations for each student after reminding them about the student's latest performance and requested them to fill in the following statements:

1. "I expect the student to work hard and improve to **achieve at least X** (out of 100%) in upcoming exams and tests."
2. "I expect the student to work hard and improve and I think that **even Y** (out of 100%) **is achievable** in upcoming exams and tests."

We also asked teachers to choose from a pre-specified list (compiled in consultation with teachers outside our study sample) three general recommendations that they thought were most important to help students improve their performance. The recommendations choice list included being more engaged in the classroom, asking questions, practicing from the textbook, practicing online, completing homework, attending virtual classrooms, and working with other students, or their parents.

¹¹ As a result, we lack statistical power to detect treatment effects using these survey-based measures. We therefore do not present these results in the paper but these are available in an online appendix.

¹² Similar to our midline and endline student survey outcomes, we lack statistical power to detect treatment effects using these survey-based teacher-reported measures and do not report them here.

2.3 Descriptive Statistics

2.3.1 Student Characteristics

We present descriptive statistics for students in our sample in Table 1. Our sample includes 1537 students from grades 3 to 8, between 6 to 15 years of age. 42% of the students are girls and 84% of the students speak Urdu, while 61% also speak English at home. We find that students are intrinsically motivated to work hard as 95% of them report wanting to get better at math. In addition, the majority of students value education highly and aspire to study for higher education. At the same time, 30% feel that they are not as good at math, and over 50% feel stressed about their current performance. Moreover, 75% of students believe that their teachers expect them to achieve over 90%. This suggests that unrealistic beliefs about what the teacher expects from them could be driving student stress. As it was a time of remote education and periodic school closures during the pandemic, engagement with teachers was also very limited: 44% report clarifying math problems with the teacher only once a week or never. Similar patterns are present concerning engagement with parents and peers. Finally, the majority of students report feeling academically motivated by their peers (79%) and report that peers do not trouble them for working hard (82%). This positive classroom environment distinguishes our setting from some other contexts that do not have conducive classroom norms such as those in [Bursztyn et al. \(2017\)](#) and [Bursztyn et al. \(2019\)](#). To corroborate this further, we measure student networks by asking students to list their friends in the classroom and find that having more friends in the classroom is positively correlated with having higher extrinsic motivation and also weakly, and positively correlated with having higher test scores.

2.3.2 Teacher Characteristics

There are 118 teachers in our sample. 59% of them have a Master's degree and are predominantly ethnically Punjabi (Table A.1.2). About 69% of teachers think that their encouragement matters the most for student performance, compared to encouragement from parents and peers. However, when asked to think about who would benefit the most from receiving high expectations and who should they be conveyed to, only 23% of the teachers reported students at the bottom end of the distribution as their choice. 30% suggest that it is best to convey expectations to those who feel most motivated. These baseline patterns motivate our intervention as teachers are aware of the importance of their expectations but do not prioritize students at the bottom end of the score distribution while thinking of conveying these expectations. These are students who can potentially have the highest marginal benefits. At the same time, teachers also acknowledge the motivational role of peers. More than 50% say that expectations should be conveyed to those who will be most successful in encouraging others, adding credence to the peer component of our study design. Finally, about a third of teachers

Table 1: Summary Statistics of Students

	Count	Mean	SD	Min	Max
Student Characteristics					
Age	1,369	10.59	1.74	6.00	15.00
Adults per Room	1,315	0.56	0.34	0.07	3.00
Female	1,537	0.41	0.49	0.00	1.00
Speaks English at home	1,468	0.64	0.48	0.00	1.00
Speaks Urdu at home	1,537	0.84	0.37	0.00	1.00
Value of Education (1-5)	1,101	4.60	0.76	1.00	5.00
Aspires to obtain Master's degree or higher	814	0.85	0.36	0.00	1.00
Classroom Engagement					
Weekly Hours doing Math Homework	1,370	2.96	4.22	0.00	30.00
Weekly Hours Studying Math	1,371	3.80	4.79	0.00	41.00
How often do you discuss math with your teacher?	1,385	1.71	0.98	0.00	3.00
How often do you discuss math with your parent?	1,385	1.79	1.10	0.00	3.00
How often do you discuss math with your peers?	1,385	0.98	0.97	0.00	3.00
Peer Characteristics					
Number of Friends in the Classroom	1,537	4.07	2.64	0.00	10.00
Stress					
Stressed about Own Performance	1,333	0.52	0.50	0.00	1.00
Stressed about Teacher's Expectations	817	0.46	0.50	0.00	1.00
Stressed about Peer's Expectations	817	0.31	0.46	0.00	1.00
Stressed about Parent's Expectations	817	0.62	0.49	0.00	1.00
Stress Index	817	0.48	0.38	0.00	1.00
Intrinsic Motivation					
Feels not good at math	1,333	0.32	0.47	0.00	1.00
Feels they work hard at math	1,333	0.87	0.33	0.00	1.00
Wants to get better at math	1,333	0.95	0.22	0.00	1.00
Intrinsic Motivation Index	1,333	0.85	0.20	0.25	1.00
Extrinsic Motivation					
Motivated by Peers	1,338	0.74	0.44	0.00	1.00
Troubled by Peers for Bad Performance	1,338	0.12	0.32	0.00	1.00
Troubled by Peers for Working Hard	1,338	0.17	0.38	0.00	1.00
Extrinsic Motivation Index	1,338	0.81	0.24	0.00	1.00

Note: The statistics are from the baseline student survey. Variables related to stress with regard to teacher's, parent's or peer's expectation, and aspirations for higher studies were only collected for the older students (in grade 5 and above) following a pilot of the survey. Students in grades 3 and 4 were asked to list up to 5 friends, while those in older grades were asked to list 10 friends. Variables measuring the number of hours doing homework or studying math exclude outliers above the 99th percentile.

report concerns about classroom disruption, attendance, or not completing homework (Table A.1.3).

3 Experiment Design

3.1 Randomization Design

Figure 1 shows the randomization design. We use a clustered randomized design at the classroom level and randomly allocate one-third of classrooms to the Expectations Arm (where student-specific high teacher expectations and encouragement are conveyed individually to a student), one-third to the Peer Arm (where in addition to conveying student-specific high teacher expectations and encouragement individually to a student, they were additionally randomly matched with another classmate and asked to encourage each other)¹³ and a one-third to a Comparison Group. Half of the Comparison Group classrooms were randomized to receive a reminder about their last test score (Information Arm) and half were randomly selected to receive no messages (Control Group). The randomization was stratified along grade¹⁴, gender composition of the school (co-educational or single gender) and the whether the average class math test score (%) in the preceding year (2020) was above or below the median. Within the Expectations Arm and Peer Arm, half the students were randomly chosen to receive the “High” teacher expectation, and half received the “Very High” teacher expectation with the corresponding statements outlined earlier (Section 2.2.3).

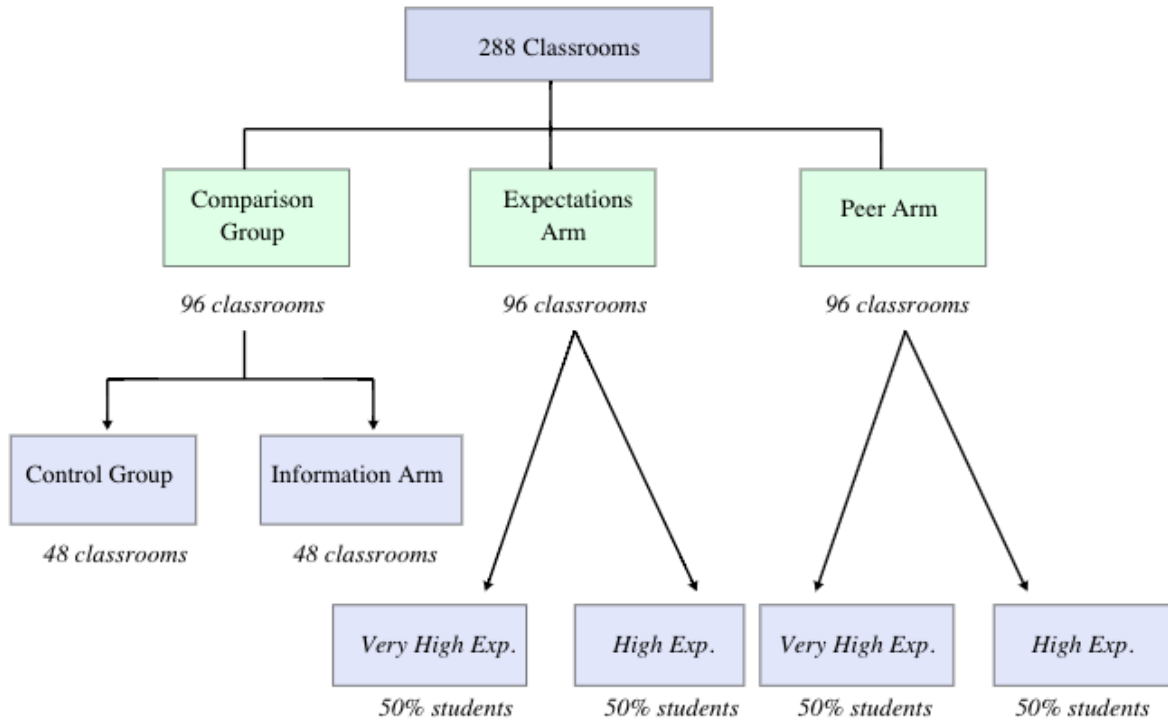
3.2 Timeline

The timeline of the study is as follows. Informed parental consent and student assent were obtained between March and May 2021. We completed the baseline teacher survey by April 2021 and the baseline student survey by May 2021. Teacher expectations were elicited and delivered by mid-June. We collected administrative test score data on student performance in exams in end-June/early July and sent a midline survey to students soon after their exams. Following the midline survey, we sent reminders to students about their teacher expectations during the start of the summer holidays and at the beginning of the new academic year in August. A final round of teacher expectations with updated design graphics was sent in November 2021 before the school conducted its end-of-term exams in December 2021. We conducted student and teacher and endline surveys between November 2021 to January 2022. A final round of

¹³In the Peer Arm, we randomly matched students with another student of the same gender, taking into account the cultural norms in the Pakistani context.

¹⁴We use a binary variable to indicate Grade 3 students (very young and unable to complete the survey without enumerator instructions and outside of class) separately from grades 4-8 (older grades)

Figure 1: Randomization Design



follow-up surveys with school administrators and students was conducted between March to May 2022.

3.3 Format of Delivery of Teacher Expectations

The delivery of teacher expectations was done via emails as the schools had switched to using virtual learning amidst physical school closures due to the pandemic. An enumerator was added to each Google classroom as a co-teacher to email students privately (which the teacher would not be able to see). Students in the Expectations Arm and Peer Arm received their most recent math test score and teacher's expectation ("High" or "Very High") according to their treatment status.¹⁵ In the Peer Arm, students first received an individual email with their test scores and their teacher's expectations and then received a joint email with their matched classmate with the additional line '*We hope you both will encourage each other*'. The joint email (and infographic) did not have any students' test scores or teachers' expectations. Students in the Information Arm received a reminder about their last score through a placebo image.

The infographic used to deliver teacher expectations signals that each student is a superhero

¹⁵For 12% of the students, teachers had reported lower expectations than the student's previous score. In these cases, students only received their teacher's expectations on the infographic and not their previous scores.

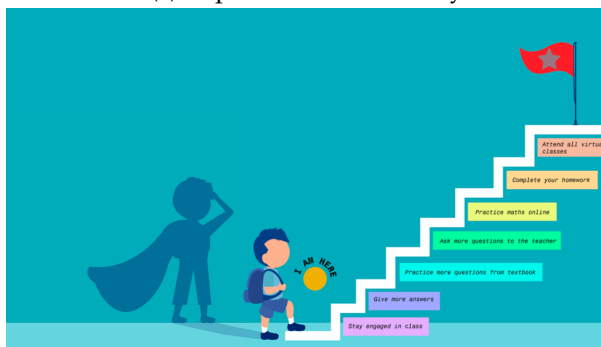
and can work towards achieving the teacher’s high expectations (using general tips given on the stairs). Figure 2 shows the designed graphics sent out during midline to each group. Appendix Figure A.2.1 illustrates the variation in “High” and “Very High” statements. Before the endline, we re-designed the graphics as shown in Appendix Figure A.2.2.

Figure 2: Treatment Delivery Design Variations - Round 1

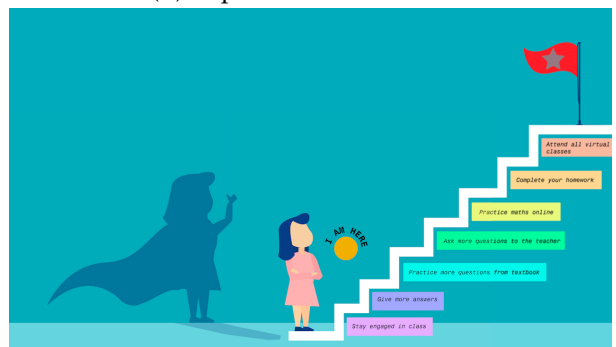
(a) Information Arm - Boy (b) Information Arm - Girl



(c) Expectations Arm - Boy



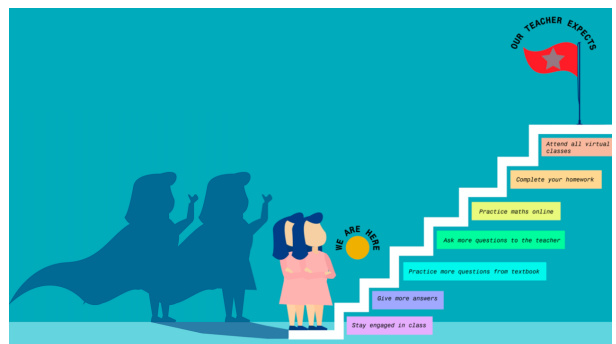
(d) Expectations Arm - Girl



(e) Peer Arm - Boy



(f) Peer Arm - Girl



4 Empirical Strategy

4.1 Empirical Specification

4.1.1 Intent to Treat Effects

Our main specification regresses pre-specified outcomes on $Information_c$, $Expectation_c$, and $Peer_c$ which equal 1 if the student is in a classroom c in the Information, Expectations, or the Peer Treatment Arms respectively.¹⁶ We use the pooled sample combining data from the midline and endline waves for the main results.¹⁷

$$Y_{ic} = \beta_0 + \beta_1 Information_c + \beta_2 Expectation_c + \beta_3 Peer_c + \phi_s + \epsilon_{ic}$$

Standard errors are clustered at the class level (unit of treatment). We include strata fixed effects ϕ_s for each strata s and control for a round fixed effect equal to 1 for the endline and 0 for the midline. We present results on standardized test scores and raw test scores, controlling for baseline student performance in a value-added specification in the latter case. We report outcomes in pre-specified groups divided into multiple families depending on whether they are capturing similar treatment effects. We also compute sharpened q values for outcomes within the same family as per the method in [Benjamini et al. \(2006\)](#) and [Anderson \(2008\)](#).

4.2 Balance

Classrooms were randomly assigned to treatment and control groups ensuring balance across baseline variables of interest collected via teacher and student surveys, along with historical administrative test score data.

4.2.1 Balance across Student and Class-Level Characteristics

We adopt two approaches to check for balance. First, we show that student characteristics are balanced across control and treatment groups for the pooled sample, midline sample, and endline sample. These include student-level characteristics such as baseline math scores, gender,

¹⁶These binary variables capture intent to treat rather than actual treatment status. However, 88% of those students who completed our midline survey reported reading the emails and the proportion is balanced across the different treatment arms so we suspect that the treatment on treated results would be slightly higher but not very different than our ITT estimates. Since we do not have this indicator for all students, we are unable to run the treatment on treated regressions.

¹⁷The differences between the treatment effect on scores across midline and endline waves are not statistically significant as shown in Table [A.5.3](#). However, results for the midline and endline waves separately are also presented in the supplementary appendix.

wealth, classroom effort in terms of hours spent studying and preparing for exams, number of friends, classroom engagement, and intrinsic and extrinsic motivation. These results are shown in Tables [A.3.1](#), [A.3.2](#) and [A.3.3](#). Next, we show balance across the treatment arms at the class level using average historical scores in Math and English, class-level variables such as class size, grade, teaching experience of the teacher, teacher-reported student engagement (motivation and interaction), disruption and warnings, absenteeism, and parental engagement. These results are shown in Table [A.3.4](#). We find that control and treatment classrooms are balanced across most characteristics. However, we will also account for any balance-related concerns in our robustness specifications where we will employ Post-Double Selection Lasso as proposed in [Belloni et al. \(2014\)](#).

4.2.2 Balance and Descriptives for Teacher Expectations

In addition to checking for balance along student and class characteristics, we also report descriptive statistics from the expectations we elicited from teachers across different student characteristics to lend credibility to our research design. We find no systematic differences in teachers' expectations by student gender, age, or wealth as shown in Figure [A.4.1](#). By reminding teachers about every student's recent math score before writing their expectations, we minimized the risk of gender or wealth-related stereotypes driving their expectations.

We also confirm that there are no systematic differences in teacher expectations across treatment and control arms (Figure [A.4.2](#)). This implies that treatment effects are unlikely to be driven by changes in teachers' behavior while writing these expectations. The results can therefore be interpreted as the effects of explicitly *communicating* the expectations to students. Additionally, as mentioned earlier, teachers were blind to the classroom treatment status as the intervention messages were delivered to students individually via email so teachers could not see the content of the message or who it was emailed to.

On average, the gap in the teacher's expectation and student performance was highest for students in the bottom 25% of the score distribution (Figure [A.4.3](#)). This suggests that teachers believed in the ability of students in the bottom quartile of the score distribution to improve substantially – even though the majority did not prioritize delivering the expectations to them as reported in the baseline survey.

5 Results

5.1 Main Results

5.1.1 Effect on Math Performance

Table 2 presents the treatment effects from our main specification on Math scores on high-stakes tests conducted by our partner schools. Column (1) reports standardized test scores and column (2) reports raw percentage scores. We find that students in the Expectations Arm score 0.19σ higher than students in the Control Group (significant at 1%). This is equivalent to a 3.3 percentage point increase in percentage scores. At the same time, we find that students who received information about their previous test scores also score 0.16σ (significant at 5%) higher than students in the Control Group, equivalent to a 2.7 percentage point increase in percentage scores. We find that the effect of the Information Arm is not statistically distinguishable from the effect of the Expectations Arm. It is important to note that these are intent-to-treat effects. While 88% of the midline survey sample reported reading the emails, the actual treatment effects are likely to be higher.

This suggests that receiving a message on behalf of the teacher that contains just a reminder about the student's past performance can also increase student performance and be just as effective as teacher expectations. In Section B.2, we present a simple effort-choice model to interpret how expectations, information, and peer comparisons can affect extrinsic motivation and achievement. We also unpack the mechanisms underlying this effect in detail in the next section and find that this is driven by schools with low parental literacy and students interpreting this reminder as an encouragement message from the teacher. Treatment effects of the Expectations and Information Arm separately for midline and endline tests are consistent with the results from the pooled specification. These results are shown in Tables A.5.1 and A.5.2. While the effects are smaller in magnitude in the endline and insignificant for the Information Arm, the differences in treatment effects over time are not significant.

Finally, we find no average effects of the Peer Arm on test scores. Further, the difference between the effects of the Expectations and Peer Arm is statistically significant. This is particularly surprising since the Peer Arm adds the peer matching component to the Expectations Arm. This finding suggests that while students may benefit from receiving teacher expectations, they may, on average, be negatively affected by being matched with a random classmate leading to an overall null effect. In the next section, we leverage the fact that peers were matched randomly to provide evidence of heterogeneous treatment effects. This will allow us to understand why the Peer Arm did not succeed in improving test scores on average.

Table 2: Treatment Effects on Test Scores

	(1) Standardized	(2) Raw
<i>Panel A. Targeted Subject: Math Scores</i>		
Expectations	0.186*** (0.066)	3.261** (1.374)
Peer	0.069 (0.074)	1.086 (1.361)
Information	0.161** (0.077)	2.747* (1.435)
<i>Comparisons (p-values)</i>		
Exp vs Peer	0.049	0.027
Exp vs Info	0.688	0.639
Info vs Peer	0.201	0.128
Observations	2773	2640
<i>Panel B. Spillover Subject: English Scores</i>		
Expectations	-0.012 (0.077)	0.261 (1.260)
Peer	-0.111 (0.086)	-1.066 (1.340)
Information	0.015 (0.089)	0.809 (1.360)
<i>Comparisons (p-values)</i>		
Exp vs Peer	0.167	0.210
Exp vs Info	0.720	0.608
Info vs Peer	0.119	0.096
Observations	2411	2413

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ The results are from pooled regressions of midline and endline scores. The scores in column (1) are standardized by using the mean and standard deviation of scores in control schools in historical administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for a student's baseline score. Regressions include strata and round fixed effects and standard errors are clustered at the level of randomization.

We also employ the Post Double Selection Lasso strategy (Belloni et al., 2014) to show that the treatment effects on test scores do not change even after accounting for any baseline characteristics that might be correlated with treatment indicators (Table A.7.1).

Alternative Specification: As the Peer Arm adds a peer matching component to the Expectations Arm, to separate the effects of expectations from the peer matching, we present the results of an alternative specification in Table A.7.3 where we define three binary variables for information, expectations, and peer matching as follows: ‘Information’ takes a value of 1 for the Information Arm, ‘Expectations’ takes a value of 1 for *both* the Expectations and Peer Arm and ‘Peer matching’ takes a value of 1 only for the Peer Arm.¹⁸ The table shows that expectations had a significantly positive effect on math scores equal to 0.19σ and peer matching had a significantly negative effect of 0.12σ (both significant at 5%). As before, we do not detect any differences between the effect of information and expectations.

5.1.2 Effect on English Performance

In addition to the above results on math test scores, Table 2 shows that none of the treatment arms have any spillover effect on English test scores in the pooled sample. However, when we separate the results at the midline and endline, we find that all three treatment arms have a positive but insignificant effect on English test scores in the midline but a negative effect of -0.18σ and -0.24σ in the endline which is significant at 10% and 5% for the Expectations and Peer Arm respectively (Tables A.5.4 and A.5.5).

5.2 Heterogeneity

We now unpack the main results further by considering heterogeneity along several important dimensions, some of which were randomized as a part of our intervention.

5.2.1 Magnitude of Expectations

First, we exploit the exogenous variation in the type of expectations delivered (i.e. ‘High’ or ‘Very High’). The results are presented in Table 3. We find that the Expectations Arm significantly raises test scores when expectations are high enough i.e., students who received a ‘Very High’ expectation from teachers, scored 0.23σ higher in math compared to the Control Group (significant at 1%). Additionally, the effect on those students who were given a ‘High’

¹⁸In our follow-up survey students reported that the scope of improvement was the most noticeable in the Expectations Arm infographic so we do not believe that it would be accurate to separate the expectations as being separate and additive to the effect of reminding students about their performance in the Expectations and Peer Arm. Hence, we keep a separate indicator for the Information Arm.

expectation is positive but not statistically significant at midline and endline. This provides evidence for the hypothesis that providing students with ambitious goals set by teachers can have high returns and does not lead to frustration. This result is similar even when we consider the midline and endline waves separately.

Panel B in Table 3 shows the results of the specification where we regress the scores on the treatment arms interacted with the gap between the student’s baseline score and the expectation delivered to them. We find that the effect of both the Expectation and Peer Arm is higher among students for whom this gap is larger. We find that a 10 percentage point increase in the gap leads to a 0.1 standard deviation increase in the impact of the Expectations and Peer Arm. This implies that receiving a higher expectation relative to one’s performance increased test scores.

5.2.2 Characteristics of the Matched Peer

Next, we exploit the random variation in matching in the Peer Arm to examine the heterogeneity of treatment effects along the characteristics of the randomly matched peers. We employ the method in Chernozhukov et al. (2018) and find evidence of substantial heterogeneity, as shown in Figure A.6.1. To systematically explore this, we first construct a measure of homophily among the matched peers as a measure of their similarity in terms of baseline characteristics such as baseline scores, teacher expectations, classroom motivation, parental wealth, and number of friends in the classroom. We construct the index by first generating the squared differences in terms of these characteristics, standardizing these differences, and then constructing an inverse variance weighted average (Anderson, 2008). The homophily index is the negative of this average.

As shown in 4, the effect of the Peer Arm is higher for whom the homophily index is higher. We find that the effect of the peer treatment arm is negative for students for whom the homophily index is low and positive for those for whom it is high. We break this down further by looking at how the treatment effect within the Peer Arm differs by the extent of similarity in terms of teacher expectations and baseline scores within matched pairs in Panels B and C of Table 4 respectively. We find that scores are significantly higher by approximately 0.3σ for individuals matched with peers who have received similar or lower teacher expectations than them when compared to those who were matched with someone who received a higher expectation than them. This is reinforced by our follow-up survey (discussed in more detail later), in which students report that they would feel disappointed and less motivated if their matched peer received a higher expectation than them. Reinforcing these patterns of heterogeneity of treatment effects of the Peer Arm, we find that the effect of being matched with a peer with the same baseline score is also 0.3σ higher than being matched with someone with a higher baseline score.

Table 3: Heterogeneity with Statement and Magnitude of Expectation Delivered

	(1) Standardized	(2) Raw
<i>Panel A. By the Type of Expectation Statement Delivered</i>		
Expectations (Very High)	0.233*** (0.070)	3.525*** (1.346)
Expectations (High)	0.123 (0.076)	2.435 (1.496)
Peer (Very High Expectation)	0.047 (0.078)	0.398 (1.320)
Peer (High Expectation)	0.090 (0.086)	1.117 (1.596)
Information	0.156** (0.075)	2.483* (1.351)
<i>Comparisons (p-values)</i>		
Exp (Very High) vs Exp (High)	0.118	0.342
Exp (Very High) vs Info	0.262	0.374
Exp (High) vs Info	0.662	0.970
Peer (Very High) vs Peer (High)	0.576	0.599
Peer (Very High) vs Info	0.167	0.074
Peer (High) vs Info	0.443	0.342
Observations	2773	2640
<i>Panel B. By the Gap between Expectation and Baseline Score</i>		
Expectations	0.103 (0.072)	1.640 (1.352)
Peer	0.053 (0.073)	0.598 (1.312)
Information	0.134* (0.081)	2.558* (1.492)
Expectations x Gap between Expectations and Baseline Score	0.013*** (0.003)	0.280*** (0.060)
Peer x Gap between Expectations and Baseline Score	0.010** (0.004)	0.198** (0.078)
Observations	2403	2403

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The results are from pooled regressions of midline and endline scores. The gap in panel B is the difference between the expectation delivered to the student and their performance. The scores in column (1) are standardized by using the mean and standard deviation of scores in control schools in historical administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for a student's baseline score. Regressions include strata and round fixed effects and standard errors are clustered at the level of randomization.

Table 4: Heterogeneity in Treatment Effects by Matched Peer Characteristics

	(1) Standardized	(2) Raw
<i>Panel A. By Homophily Index (Whole Sample)</i>		
Expectations	0.148*** (0.052)	2.860*** (1.018)
Information	0.125* (0.065)	2.538** (1.213)
Peer	-0.742** (0.295)	-15.317*** (5.542)
Peer x Homophily Index	1.081*** (0.348)	21.910*** (6.502)
Observations	2512	2512
<i>Panel B. By Matched Peer's Expectation (Within Peer-Arm)</i>		
Own expectation	0.029*** (0.006)	0.581*** (0.103)
Peer's expectation is same	0.354*** (0.119)	6.312*** (2.192)
Peer's expectation is lower	0.295** (0.121)	5.131** (2.251)
Constant	-2.592*** (0.498)	25.501*** (8.363)
Observations	554	554
<i>Panel C. By Matched Peer's Baseline Score (Within Peer-Arm)</i>		
Own score	0.021*** (0.006)	0.416*** (0.099)
Peer's score is same	0.310** (0.144)	5.799** (2.572)
Peer's score is lower	0.047 (0.142)	0.673 (2.582)
Constant	-1.700*** (0.432)	44.049*** (7.464)
Observations	552	552

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The results in Panel A are from pooled regressions of midline and endline scores. The Homophily Index is a measure of the similarity between matched peers in terms of baseline characteristics such as scores, teacher expectations, classroom motivation, parental wealth, and number of friends in the classroom. Panel B and Panel C show within Peer-Arm regression results. The scores in column (1) are standardized by using the mean and standard deviation of scores in control schools in historical administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for a student's baseline score. Regressions include strata and round fixed effects and standard errors are clustered at the level of randomization.

When compared with the Control Group, Appendix Table A.6.1 shows that students who were matched with a peer with the same baseline score achieve a 0.17σ higher test score (significant at 10%) than the control group. This effect is not statistically distinguishable from that of the Expectations Arm. However, being matched with a peer with a higher baseline score does not improve student performance. Even though this effect is not statistically distinguishable from that of being matched with someone with the same score, we find that it is significantly lower than the effect of the Expectations Arm. This reinforces the finding that peers who are similar in terms of baseline scores perform significantly better than the Control Group and have a treatment effect as large as those who were in the Expectations Arm.

5.2.3 Score Distribution

First, we run quantile regressions and show that the treatment effects of the Expectations and Information Arms discussed above are driven by positive effects on students at the bottom and middle of the distribution of baseline math test scores. Figure A.6.1 plots the treatment effects on different quantiles of the score distribution. We find that the treatment effects of the Expectations Arm and the Peer Arm are higher for lower quantiles of performance and decline as the score increases. The Peer Arm has no effect on average and displays little heterogeneity across the quantiles of the baseline student test score distribution.

The positive effect on this subgroup is further validated in Table A.6.2 where we employ the strategy recommended in Abadie et al. (2018). This strategy predicts math performance for the control group using a set of covariates selected by LASSO from a list including variables measuring demographic characteristics, classroom engagement, academic effort, and motivation. We de-bias the prediction process and deal with “endogenous stratification” by computing the leave-one-out and repeated split sample estimators. This model is then used to predict performance for all students and classify them into four subgroups for which we separately compute heterogeneous treatment effects. These results are shown in Table A.6.2 where we find evidence that the treatment effects are strongest for students who were predicted to perform poorly. For instance, we find with the repeated split sample estimation that the effect of the Expectations Arm on test scores of the students predicted to perform the worst is 0.33σ and significant at 1%. By contrast, the effect on those predicted to perform the best is 0.004σ and not statistically significant. The effect of the Information Arm is also concentrated among these subgroups. The Peer Arm has no effects across any of these groups.

6 Discussion

6.1 Mechanisms

We use school administrative data and school-related information shared via a school survey with head teachers, along with our findings from heterogeneity analysis, and the follow-up student survey to inform our understanding of the mechanisms behind the treatment effects in each treatment arm.¹⁹

6.1.1 Expectations Arm

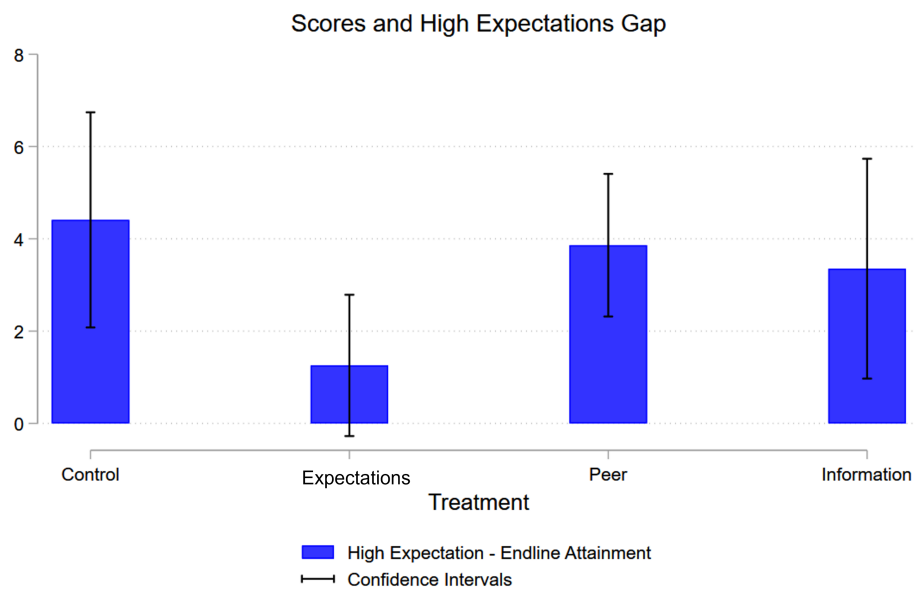
The infographic delivered to students in the Expectations Arm contained information about their current performance, their teacher's expectations, and generalized tips that they can follow to achieve them. We believe that the treatment effect is driven by teacher expectations rather than the information about the current performance or the tips provided to the students for the following reasons.

First, our follow-up survey reveals that students primarily notice the potential for improvement and the expectations set by their teacher, rather than their actual score (Figure A.8.1).²⁰ Second, students overwhelmingly (80%) interpreted the Expectations Arm image as a goal-setting mechanism and a form of encouragement from their teacher, rather than as a commentary on how smart they are (Figure A.8.2a). Consistent with this, we find that the gap between teacher expectations and the score achieved at the endline is the smallest for students in the Expectations Arm. In particular, Figure 3 shows that this gap was statistically indistinguishable from zero for students who received the 'High Expectations' statement ("... achieve at least"). Similarly, the gap was smaller in magnitude (6 percentage points) for students who received the 'Very High Expectations' statement ("even... is achievable") than students in the Peer Arm and Information Arm (9 percentage points) although we are not statistically powered to show that these differences are significant.

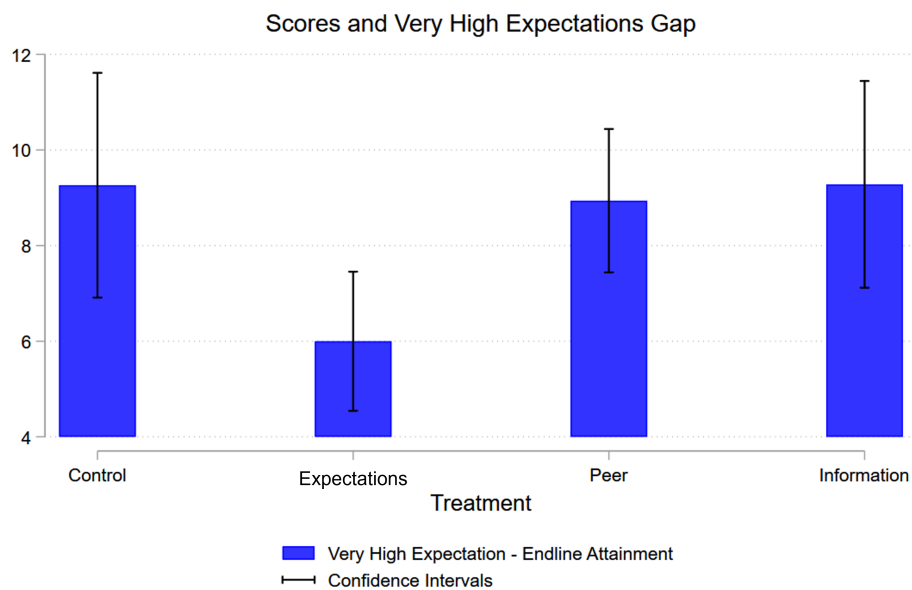
¹⁹Since the midline and endline survey completion rates were less than 50%, our power in detecting effects on self-reported student outcomes (such as self-efficacy, effort, motivation etc) listed in our pre-analysis plan is limited. We do not detect any treatment effects and results are available upon request.

²⁰In fact, only 15% of the students reported that they would notice their score the most in the image.

Figure 3: Gap between teacher expectations and student performance at endline



(a) Gap between “High” Expectations and Score Across Treatment Arms.



(b) Gap between “Very High” Expectations and Score Across Treatment Arms.

Note: Panel (a) plots the gap between ‘High’ expectations elicited from teachers and students’ endline score across the treatment arms with 95% confidence intervals. Panel (b) plots the gap between ‘Very High’ expectations elicited from teachers and students’ endline scores across the treatment arms with 95% confidence intervals.

Furthermore, we also find that the magnitude of the expectation affects the impact of this treatment arm. In particular, a 10 percentage point increase in the gap between expectation and baseline performance leads to a 0.1 standard deviation increase in the impact of the Expectations Arm (Table 3). Finally, the tips on the infographic are unlikely to be new information to students as all the teachers unanimously reported that they were already conveying tips to students about how they can improve in our baseline survey. Moreover, the tips on the infographic were not student-specific and very generic (e.g. 'Being more engaged in the classroom, Completing homework, etc.). The effectiveness of the Expectations Arm therefore lies in its ability to instill a sense of improvement and motivation among students, rather than simply providing informational content.

6.1.2 Information Arm

The delivery of students' previous test scores via the placebo image likely did not provide new information, as students typically receive report cards with their scores at the end of each term. Nearly 80% of the schools in our sample send out report cards at the end of each term. In fact, for older grades (5-8), 20% of the schools send out report cards every month, and 7% do so for younger grades (3-4). Since the first round of our intervention was delivered close to the end of the term and the second round after the end of the term, the treatment effect is unlikely to be driven by pure information effects.

Further, although the delivery of report cards as a hard copy does vary among schools, our analysis found no significant differences in treatment effects between 44% of schools that send printed report cards home (in addition to SMS and online links) versus those that do not (Table A.6.3). Expressing scores as a percentage (out of 100%) used in our intervention aligns with the common practices in schools. 69% of the schools in our sample give out scores in percentages and 88% of schools use raw scores. Therefore, it is unlikely that the treatment effects observed in this arm can be attributed to the novelty or format of the information delivered.

However, we posit that the ease of interpreting information could be a critical factor driving the observed positive treatment effects in this arm. 75% of the schools reported that parents have medium literacy levels and 13% mentioned low literacy levels. Additionally, 88% of the schools reported limited technological proficiency among parents. Moreover, Table 5 shows significantly greater effects of information in schools with low parental literacy compared to schools with high parental literacy. Importantly, we do not observe this heterogeneity in the Expectations Arm. Further, 30% of the students in our follow-up survey reported that they do not recall their scores, and the image is a helpful reminder and over 50% of students reported that even though they remember their scores, they still find the image helpful as a reminder. Therefore, delivering information in a simple, visually accessible format in the Information Arm could enhance comprehension and retention.

Table 5: Heterogeneity in Treatment Effects by Parental Literacy

	(1) Standardised Score	(2) Percentage Score
Expectations	0.035 (0.110)	0.711 (1.785)
Peer	-0.334*** (0.122)	-3.990* (2.331)
Information	-0.373*** (0.102)	-6.277* (3.354)
Low	-0.482*** (0.155)	-1.744 (2.682)
Medium	-0.397*** (0.118)	-3.777* (2.054)
Expectations x Low Literacy	0.077 (0.181)	0.051 (3.598)
Information x Low Literacy	0.673*** (0.222)	8.893* (4.767)
Peer x Low Literacy	0.344* (0.204)	1.586 (3.775)
Expectations x Medium Literacy	0.178 (0.136)	3.450 (2.443)
Information x Medium Literacy	0.561*** (0.137)	9.934*** (3.721)
Peer x Medium Literacy	0.448*** (0.150)	6.272** (2.871)
Constant	0.257** (0.125)	43.564*** (3.833)
<i>Comparisons (p-values)</i>		
Treatment Effect (High Literacy - Low Literacy): Info v/s Exp	0.003	0.066
Treatment Effect (High Literacy - Medium Literacy): Info v/s Exp	0.004	0.088
Treatment Effect: Info v/s Exp (Low Literacy)	0.252	0.560
Treatment Effect: Info v/s Exp (Medium Literacy)	0.730	0.667
Treatment Effect: Info v/s Exp (High Literacy)	0.000	0.052
Observations	2773	2640

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ The results are from pooled regressions of midline and endline scores. The scores in column (1) are standardized by using the mean and standard deviation of scores in control schools in historical administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for a student's baseline score. Regressions include strata and round fixed effects and standard errors are clustered at the level of randomization.

Finally, the majority of students reported feeling encouraged or motivated upon receiving the image. Nearly 40% of the students inferred that the teacher’s intention for sending this image and information was to encourage them (Figure A.8.2a). This underscores the potential of delivering performance-related information in a targeted and encouraging manner to positively influence student perceptions and ultimately, their academic performance. One caveat here is that it matters whether the students think they were the only ones receiving the image versus if the entire class was receiving it. In particular, while most students still inferred that the teacher was trying to encourage them when sending a reminder about their last score, students commonly expressed that they would feel indifferent or no reaction if the image was sent to all (Figure A.8.2b). This indicates that the belief that it is a targeted reminder from the teacher plays a crucial role in its effectiveness.

6.1.3 Peer Arm

While the Expectations Arm had large positive and significant treatment effects, we find that additionally pairing two classmates randomly resulted in an overall treatment effect statistically indistinguishable from zero. This was not due to unfavorable classroom norms. In fact, in our follow-up student survey, 62% of the students reported that they would be more motivated and happier when paired with another classmate and asked to encourage one another, and an additional 10% mentioned that they would be less stressed (Figure A.8.3b). Students were also unlikely to see the matched peer as a competitor and inferred that they were matched to help each other with nearly 40% of students thinking that the teacher was encouraging them to study together and improve together (Figure A.8.3a).

However, we believe morale effects due to relative comparisons between matched students are likely driving the heterogeneity of treatment effects in this arm. In particular, the majority of the students reported that they would feel disappointed or sad when matched with a peer with a higher teacher expectation. By contrast, when asked how they would feel if they were paired with a similar-scoring peer or a peer with similar teacher expectations, students reported they would feel happy and motivated. Interestingly, when matched with a peer with lower achievement, around one-third of students reported they would feel sad or disappointed, and a third reported they would feel indifferent. This also aligns well with the heterogeneity of treatment effects we document when the matched peer was similar or different in terms of several characteristics including test scores and teacher expectations. We also find that when asked what the students would do following being randomly paired, around one-third of students reported that they would try to find out what their peer scored and the teacher’s expectations for them. Based on these findings on student reactions to being paired, we believe that the morale effects resulting from relative comparisons likely explain the null average treatment effect.

6.2 Cost-benefit analysis

This intervention yields an incredibly low-cost way to boost student performance. In particular, designing the infographics amounted to \$0.17 per student (Appendix Table A.9.1) in our study. We did not incur any additional costs in delivering expectations to students and we do not anticipate any additional costs for schools since these expectations can be easily delivered in the classroom or included in report cards. Given that the treatment effect size was 0.19 and 0.16 standard deviations in the Expectations Arm and Information Arm respectively, this implies that a 0.1σ increase in test scores costs 9 cents per student in the Expectations and 10 cents per student in the Information Arm. For reference, this is orders of magnitude smaller than several interventions that have been implemented to raise test scores in developing countries (Glewwe and Muralidharan, 2016). For example, Blimpo (2014) monetary performance-based incentives for students had a cost of \$1 – 3 per 0.1σ increase in student test scores in Benin, and performance-pay based teacher incentives cost \$1 per 0.1σ increase in student test scores in India (Muralidharan and Sundararaman, 2011). Furthermore, this intervention is also easily sustainable by schools without the researchers’ involvement as it leverages information that is internal to the school. The cost-effectiveness and simplicity of the intervention enhance the scalability and adoption possibility of this practice by school systems.

7 Conclusion

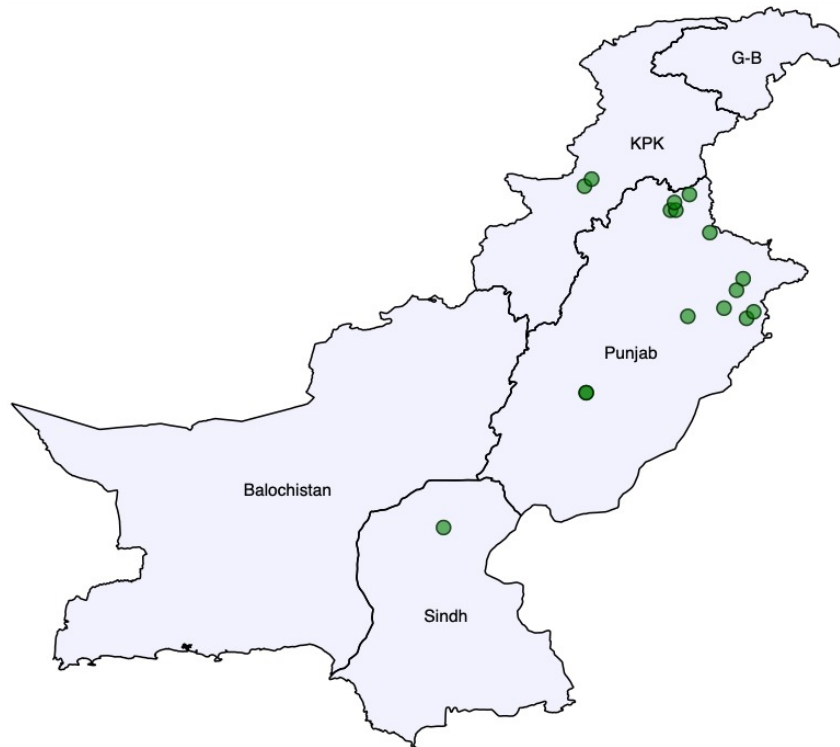
Our paper investigates the effect of conveying student-specific teacher expectations of high effort and achievement on student performance. We find that students respond positively to ambitious expectations set and communicated by their teachers. The results are more pronounced for students at the lowest end of the achievement distribution. We also find that in addition to communicating expectations, nudging peers who are similar in attributes such as academic performance and teachers’ expectations to encourage each other can further improve student achievement. Finally, we find that teachers giving reminders about past performance to students also improves their scores in a setting with low parental literacy, with effects as large as the effects of expectations.

These results have implications for policy and practice in schools. In contrast to other resource and time-intensive interventions that have been implemented to address the ‘learning crisis’ in resource-constrained settings, we show that providing tailored and realistic teacher’s expectations to students in an easy-to-access infographic is a light-touch and cost-effective way to improve academic performance. An exciting next step would be to understand the relationship between teachers’ expectations and parental involvement in their children’s education.

A Appendix

A.1 Context

Figure A.1.1: Geographic Locations of Schools in our Study.



Note: The map is generated using coordinates from the Stanford Geo Data Repository. The colored dots represent the schools in our sample with randomly jittered coordinates. We use acronyms for Khyber Paktunkhwa (KPK) and Gilgit-Baltistan (G-B) due to space constraints.

Table A.1.1: Summary Statistics of Schools

	Count	Mean	SD	Min	Max
Number of Yearly Parental meetings	15	2.67	0.70	2.00	4.00
How do students receive information about their performance					
Printed Report Card (Yes/No)	16	0.44	0.51	0.00	1.00
Raw Scores (Yes/No)	16	0.88	0.34	0.00	1.00
Percentage Score (Yes/No)	16	0.69	0.48	0.00	1.00
Parental literacy					
High	16	0.12	0.34	0.00	1.00
Low	16	0.12	0.34	0.00	1.00
Medium	16	0.75	0.45	0.00	1.00
Parental Economic Status					
High Income	16	0.19	0.40	0.00	1.00
Middle Income	16	0.38	0.50	0.00	1.00
Upper Middle Income	16	0.44	0.51	0.00	1.00
How comfortable are parents with technology					
Not Comfortable	16	0.12	0.34	0.00	1.00
Somewhat Comfortable	16	0.88	0.34	0.00	1.00

Note: The statistics are from the school-level head-teacher survey from 15 schools (one school had two different branches with separate school heads).

Table A.1.2: Summary Statistics of Teachers

	Count	Mean	SD	Min	Max
Teacher Characteristics					
Age	118	36.54	7.54	23.00	60.80
Number of years of experience in school	118	6.74	5.73	0.00	27.50
Ethnicity					
Punjabi	110	0.84	0.37	0.00	1.00
Sindhi	110	0.02	0.13	0.00	1.00
Pashtun	110	0.03	0.16	0.00	1.00
Other	110	0.12	0.32	0.00	1.00
Education					
Doctorate	118	0.01	0.09	0.00	1.00
Masters (M. Ed, etc)	118	0.59	0.49	0.00	1.00
Undergraduate (B. Ed, etc)	118	0.11	0.31	0.00	1.00
Highschool Graduate	118	0.04	0.20	0.00	1.00
Other	118	0.25	0.43	0.00	1.00
Who will benefit most from communication of expectations?					
Top of achievement distribution	97	0.52	0.50	0.00	1.00
Middle of achievement distribution	99	0.35	0.48	0.00	1.00
Bottom of achievement distribution	94	0.23	0.43	0.00	1.00
Whose encouragement matters the most?					
Teachers	115	0.69	0.47	0.00	1.00
Friends	98	0.10	0.30	0.00	1.00
Parents	95	0.22	0.42	0.00	1.00
Teacher Beliefs Agree/Strongly Agree with					
Students from less privileged backgrounds are less likely to succeed in math	118	0.04	0.20	0.00	1.00
Students with more educated parents are more likely to succeed in math	118	0.58	0.50	0.00	1.00
Student ability is more important than hard work to do well in math	118	0.48	0.50	0.00	1.00
Girls are better at math than boys	118	0.19	0.39	0.00	1.00
Motivation and self confidence matter more than academic performance	118	0.82	0.38	0.00	1.00
Students care about what their friends think about them	118	0.86	0.35	0.00	1.00
Working hard is not considered cool among students	118	0.38	0.49	0.00	1.00

Note: The statistics are from the baseline teacher survey. We asked teachers to rank from 1-3 who they thought would benefit the most from the communication of teacher expectations, e.g., 52% of teachers ranked the top of the achievement distribution as 1.

Table A.1.3: Summary Statistics of Classes

	Count	Mean	SD	Min	Max
Classroom Characteristics					
Class size	282	20.90	4.76	7.00	34.00
Teacher taught class for > 1 year	288	0.59	0.49	0.00	1.00
Teacher's Perception of Class					
Class is interactive	252	0.48	0.50	0.00	1.00
Class is motivated	252	0.39	0.49	0.00	1.00
Class is disruptive	252	0.02	0.14	0.00	1.00
Teacher gave warnings for disruption	252	0.28	0.45	0.00	1.00
Teacher gave warnings for homework	252	0.35	0.48	0.00	1.00
Teacher gave warnings for attendance	252	0.37	0.48	0.00	1.00
Percentage of students absent in last math class	245	17.99	17.77	0.00	80.00
Overall parental interest	251	0.41	0.49	0.00	1.00

Note: The statistics are from the baseline teacher survey. For each of the classes taught by a teacher, we elicited information about student behavior in those classes.

A.2 Treatment Delivery Variations

Figure A.2.1: Treatment Delivery Illustrations - Round 1

(a) Illustration for Student-Specific “High” Teacher Expectation - Girl



(b) Illustration for Student-Specific “Very High” Teacher Expectation - Girl

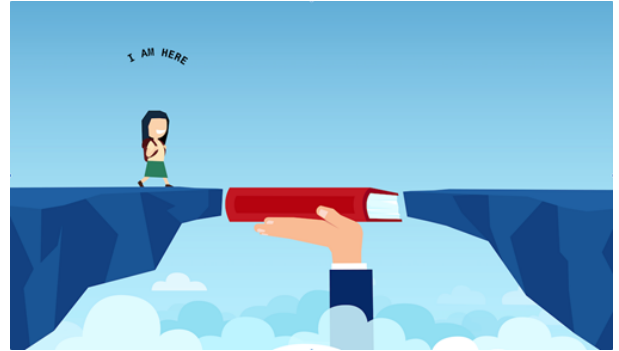


Figure A.2.2: Treatment Delivery Variations- Round 2

(a) Control Group (with Score) - Boy



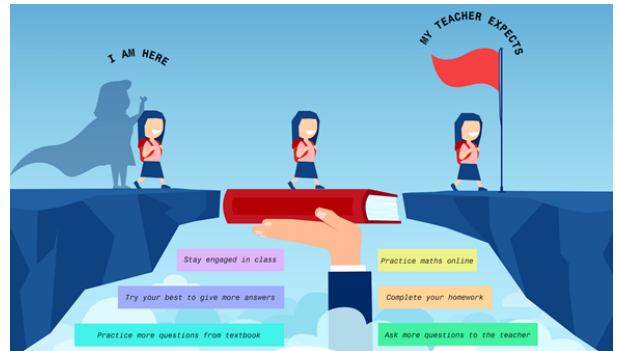
(b) Control Group (with Score) - Girl



(c) Individual Arm - Boy



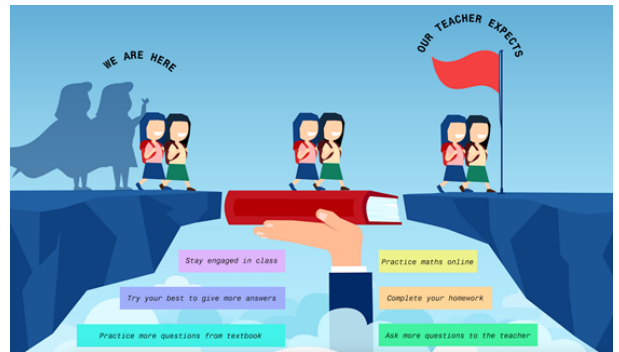
(d) Individual Arm - Girl



(e) Peer Arm - Boy



(f) Peer Arm - Girl



A.3 Balance Tables

Table A.3.1: Balance Table of Student Characteristics (Pooled Student Scores Sample)

	Mean				P-values		
	(1) Control	(2) Exp	(3) Peer	(4) Info	(1)-(2)	(1)-(3)	(1)-(4)
Baseline Math Score	82.78	83.01	83.18	85.13	0.69	0.99	0.23
Female	0.43	0.48	0.33	0.38	0.01***	0.01**	0.60
High Parental Income	0.15	0.13	0.08	0.12	0.69	0.13	0.80
Adults per Room	0.57	0.55	0.56	0.58	0.41	0.84	0.26
High Parental Literacy	0.10	0.11	0.09	0.06	0.69	0.93	0.90
Number of Friends in the Classroom	4.21	4.15	3.92	4.03	0.42	0.10*	0.95
Weekly Hours Studying Math	4.08	3.81	3.91	3.46	0.90	0.84	0.17
Weekly Hours doing Math Homework	3.29	2.73	3.23	2.54	0.14	0.13	0.09*
Teacher Takes Interest in Studies	0.96	0.95	0.95	0.95	0.95	0.71	0.89
How often do you discuss math with your teacher?	1.71	1.70	1.70	1.74	0.83	0.81	0.71
How often do you discuss math with your parent?	1.83	1.82	1.71	1.98	0.58	0.04**	0.02**
How often do you discuss math with your peers?	0.92	1.01	1.01	0.84	0.35	0.35	0.02**
Intrinsic Motivation Index	0.85	0.84	0.85	0.86	0.39	0.75	0.14
Extrinsic Motivation Index	0.84	0.82	0.80	0.79	0.31	0.15	0.16
Observations:	507	966	914	431			

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Midline student scores sample is used to check for balance on baseline student characteristics. Columns 1-4 report the averages for the four comparison groups. The next three columns report p-values from the regression of baseline characteristics on the treatment dummy. The column heading indicates the comparison, e.g., (1)-(2) reports the difference between the expectations arm and the control group and whether or not the difference is statistically significant. The regression controls for strata fixed effects and is clustered at the classroom level. The variables 'High Parental Literacy' and 'High Parental Income' capture the school heads' report on whether parents in their school have high literacy and income (i.e. these measures were not collected at the student level).

Table A.3.2: Balance Table of Student Characteristics (Midline Student Scores Sample)

	Mean				P-values		
	(1) Control	(2) Exp	(3) Peer	(4) Info	(1)-(2)	(1)-(3)	(1)-(4)
Baseline Math Score	82.93	82.92	83.33	84.77	0.64	0.89	0.34
Female	0.44	0.48	0.34	0.38	0.01**	0.01**	0.45
High Parental Income	0.15	0.14	0.09	0.13	0.77	0.18	0.74
Adults peer Room	0.57	0.55	0.55	0.59	0.55	0.60	0.14
High Parental Literacy	0.10	0.11	0.10	0.06	0.77	0.76	0.78
Number of Friends in the Classroom	4.16	4.19	3.90	4.06	0.25	0.05*	0.93
Weekly Hours Studying Math	4.04	3.74	3.88	3.45	0.82	0.79	0.18
Weekly Hours doing Math Homework	3.27	2.68	3.25	2.61	0.08*	0.10	0.18
Teacher Takes Interest in Studies	0.96	0.95	0.95	0.96	0.99	0.86	0.94
How often do you discuss math with your teacher?	1.70	1.70	1.73	1.73	0.75	0.71	0.88
How often do you discuss math with your parent?	1.81	1.80	1.69	1.96	0.59	0.04**	0.02**
How often do you discuss math with your peers?	0.94	1.02	1.01	0.85	0.37	0.37	0.02**
Intrinsic Motivation Index	0.85	0.84	0.84	0.86	0.37	0.93	0.11
Extrinsic Motivation Index	0.84	0.83	0.80	0.79	0.31	0.19	0.19
Observations:	273	532	503	229			

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Midline student scores sample is used to check for balance on baseline student characteristics. Columns 1-4 report the averages for the four comparison groups. The next three columns report p-values from the regression of baseline characteristics on the treatment dummy. The column heading indicates the comparison, e.g., (1)-(2) reports the difference between the expectations arm and the control group and whether or not the difference is statistically significant. The regression controls for strata fixed effects and is clustered at the classroom level. The variables 'High Parental Literacy' and 'High Parental Income' capture the school heads' report on whether parents in their school have high literacy and income (i.e. these measures were not collected at the student level).

Table A.3.3: Balance Table of Student Characteristics (Endline Student Scores Sample)

	Mean				P-values		
	(1) Control	(2) Exp	(3) Peer	(4) Info	(1)-(2)	(1)-(3)	(1)-(4)
Baseline Math Score	82.60	83.13	83.00	85.53	0.75	0.86	0.16
Female	0.42	0.48	0.32	0.39	0.01***	0.01**	0.81
High Parental Income	0.15	0.13	0.07	0.10	0.60	0.08*	0.87
Adults peer Room	0.57	0.55	0.56	0.58	0.28	0.84	0.47
High Parental Literacy	0.10	0.11	0.08	0.06	0.59	0.81	0.94
Number of Friends in the Classroom	4.26	4.09	3.94	3.99	0.69	0.19	0.83
Weekly Hours Studying Math	4.12	3.89	3.94	3.47	0.99	0.91	0.17
Weekly Hours doing Math Homework	3.30	2.78	3.21	2.46	0.29	0.22	0.05*
Teacher Takes Interest in Studies	0.96	0.95	0.94	0.95	0.92	0.58	0.73
How often do you discuss math with your teacher?	1.73	1.69	1.66	1.76	0.89	0.40	0.55
How often do you discuss math with your parent?	1.84	1.84	1.72	1.99	0.59	0.04**	0.02**
How often do you discuss math with your peers?	0.90	1.01	1.02	0.83	0.34	0.35	0.03**
Intrinsic Motivation Index	0.86	0.85	0.85	0.87	0.43	0.59	0.21
Extrinsic Motivation Index	0.84	0.82	0.79	0.78	0.36	0.13	0.16
Observations:	234	434	411	202			

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Endline student scores sample is used to check for balance on baseline student characteristics. Columns 1-4 report the averages for the four comparison groups. The next three columns report p-values from the regression of baseline characteristics on the treatment dummy. The column heading indicates the comparison, e.g., (1)-(2) reports the difference between the expectations arm and the control group and whether or not the difference is statistically significant. The regression controls for strata fixed effects and is clustered at the classroom level. The variables 'High Parental Literacy' and 'High Parental Income' capture the school heads' report on whether parents in their school have high literacy and income (i.e. these measures were not collected at the student level).

Table A.3.4: Balance Table of Class Characteristics

	(1)	Mean (2)	(3)	(4)	P-values		
	Control	Exp	Peer	Info	(1)-(2)	(1)-(3)	(1)-(4)
Historical Scores:							
Math Percentage	77.31	79.24	78.02	76.85	0.13	0.87	0.27
English Percentage	76.50	78.75	79.04	77.71	0.75	0.62	0.78
Class Characteristics:							
Class Size	21.04	20.74	20.55	21.79	0.68	0.39	0.16
Number of students in grade 3	0.19	0.16	0.13	0.17	0.94	0.36	0.73
Number of students in grade 4	0.21	0.16	0.17	0.19	0.56	0.84	0.74
Number of students in grade 5	0.13	0.17	0.20	0.11	0.85	0.20	0.27
Number of students in grade 6	0.15	0.18	0.23	0.19	0.64	0.24	0.99
Number of students in grade 7	0.13	0.19	0.18	0.19	0.69	0.90	0.74
Number of students in grade 8	0.21	0.16	0.09	0.15	0.66	0.10*	0.91
Teacher Evaluated Student Measures:							
Taught Class for > 1 year	0.55	0.60	0.60	0.57	0.79	0.75	0.85
Interactive	0.45	0.52	0.48	0.44	0.39	0.92	0.55
Motivated	0.36	0.44	0.37	0.38	0.28	0.54	0.91
Disruptive	0.07	0.01	0.01	0.00	0.47	0.55	0.34
Warnings for Disruption	0.24	0.29	0.29	0.28	0.79	0.79	1.00
Warnings for Homework	0.38	0.33	0.35	0.38	0.50	0.99	0.66
Warnings for Attendance	0.45	0.36	0.34	0.36	0.82	0.53	0.89
Parental Interest	0.45	0.45	0.36	0.38	0.35	0.25	0.72
Classes:	49	96	95	48			

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Historical scores are computed using the administrative data on the most recent test score (averaged at the class level and reported as a percentage) in the academic year preceding the baseline. Reports on the level of interaction, motivation and disruption, as well as warnings issued and level of parental interest, were collected from teachers for each of their classes. Columns 1-4 report the averages for the four comparison groups. The next three columns report p-values from the regression of baseline characteristics on the treatment dummy. The column heading indicates the comparison, e.g., (1)-(2) reports the difference between the expectations arm and the control group and whether or not the difference is statistically significant. The regression controls for strata fixed effects and is clustered at the classroom level.

A.4 Teacher Expectations



Figure A.4.1: Teacher Expectations Balance by Student Gender, Cohort, and Wealth Index

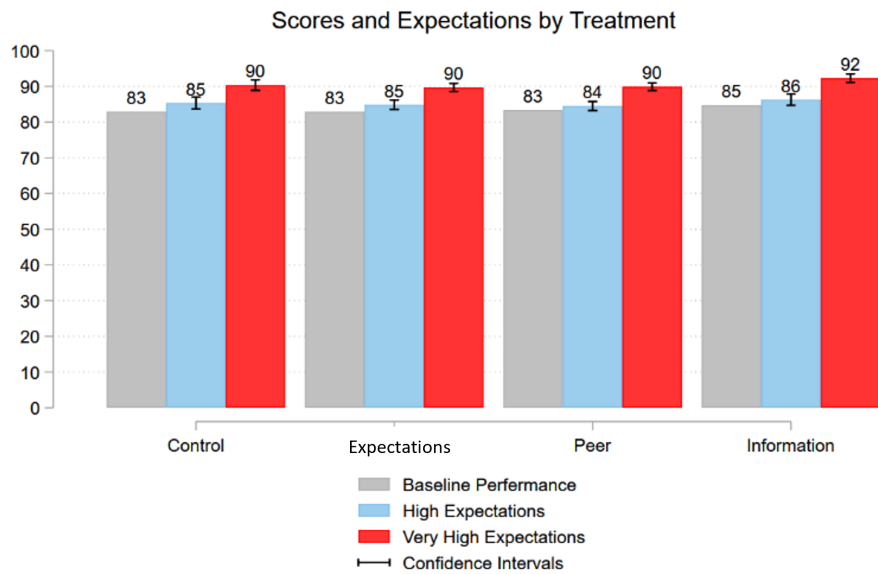
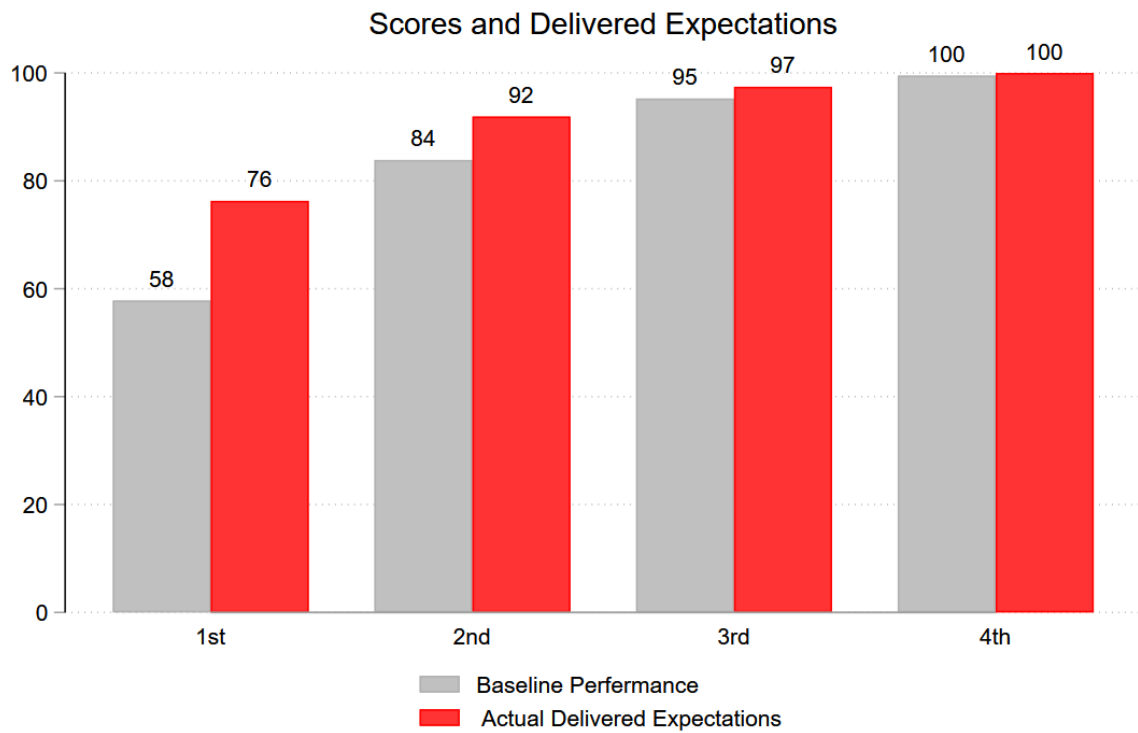


Figure A.4.2: Teacher Expectations by Treatment Arms

Note: Panel (a) plots students' baseline math scores and the randomly delivered ('High' or 'Very High') teacher's expectations of them across student gender, older (grades 6-8) or younger (grades 3-5), and wealth index. Panel (b) shows the balance in teacher expectations across the treatment arms.

Figure A.4.3: Raw scores and Teacher Expectations Delivered



Note: The figure plots students baseline math scores and the randomly delivered ('High' or 'Very High') teacher's expectations to them across four quartiles of baseline performance i.e., 1st refers to the students in the 25th percentile of baseline scores.

A.5 Midline and Endline Results (Separately)

Table A.5.1: Treatment Effects on Midline Math Test Scores

	(1) Standardised Scores	(2) Raw Scores
Expectations	0.203** (0.096)	3.751** (1.896)
Peer	0.111 (0.100)	1.757 (1.834)
Information	0.208* (0.109)	3.798** (1.864)
Constant	-0.200 (0.164)	34.112*** (4.628)
Exp vs Peer	0.233	0.156
Exp vs Info	0.952	0.975
Info vs Peer	0.282	0.136
Observations	1492	1422

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The scores in column (1) are standardized using the mean and standard deviation of the historical scores in control schools from administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for the student's baseline score. Regressions include strata fixed effects and standard errors are clustered at the level of randomization.

Table A.5.2: Treatment Effects on Endline Math Test Scores

	(1) Standardised Scores	(2) Raw Scores
Expectations	0.170* (0.093)	2.684 (1.796)
Peer	0.022 (0.102)	0.285 (1.902)
Information	0.111 (0.105)	1.623 (2.058)
Constant	0.072 (0.140)	49.829*** (4.314)
Exp vs Peer	0.050	0.061
Exp vs Info	0.467	0.496
Info vs Peer	0.323	0.403
Observations	1281	1218
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The scores in column (1) are standardized using the mean and standard deviation of the historical scores in control schools from administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for the student's baseline score. Regressions include strata fixed effects and standard errors are clustered at the level of randomization.

Table A.5.3: Treatment Effects on Math Test Scores Over Time

	(1) Standardised	(2) Raw
Information	0.184* (0.110)	3.437 (2.090)
Information x Endline	-0.048 (0.156)	-0.737 (2.957)
Expectations	0.186* (0.099)	3.358* (1.912)
Expectations x Endline	0.000 (0.144)	0.225 (2.707)
Peer	0.083 (0.103)	1.150 (1.967)
Peer x Endline	-0.030 (0.146)	-0.225 (2.727)
Endline	0.109 (0.127)	2.178 (2.379)
Constant	-0.120 (0.124)	75.416*** (2.524)
Observations	2773	2773
Info Effect (End-Mid)= Exp Effect (End-Mid)	0.666	0.661
Peer Effect (End-Mid)= Exp Effect (End-Mid)	0.763	0.811
Peer Effect (End-Mid)= Info Effect (End-Mid)	0.870	0.816
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The scores in column (1) are standardized using the mean and standard deviation of the historical scores in control schools from administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for the student's baseline score. Regressions include strata fixed effects and standard errors are clustered at the level of randomization. Endline is a binary variable equal to 1 for the scores collected during the endline round and 0 for the midline round. The t-tests reported below the table, labelled "End-Mid", check if the change in the effect of the treatment arms is differential across arms.

Table A.5.4: Treatment Effects on Midline English Test Scores

	(1) Standardised Scores	(2) Raw Scores
Expectations	0.137 (0.084)	2.518* (1.354)
Peer	0.009 (0.102)	0.676 (1.614)
Information	0.086 (0.103)	1.945 (1.545)
Constant	-0.080 (0.174)	77.117*** (2.987)
Exp vs Peer	0.153	0.171
Exp vs Info	0.591	0.684
Info vs Peer	0.473	0.413
Observations	1189	1189
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The scores in column (1) are standardized using the mean and standard deviation of the historical scores in control schools from administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for the student's baseline score. Regressions include strata fixed effects and standard errors are clustered at the level of randomization.

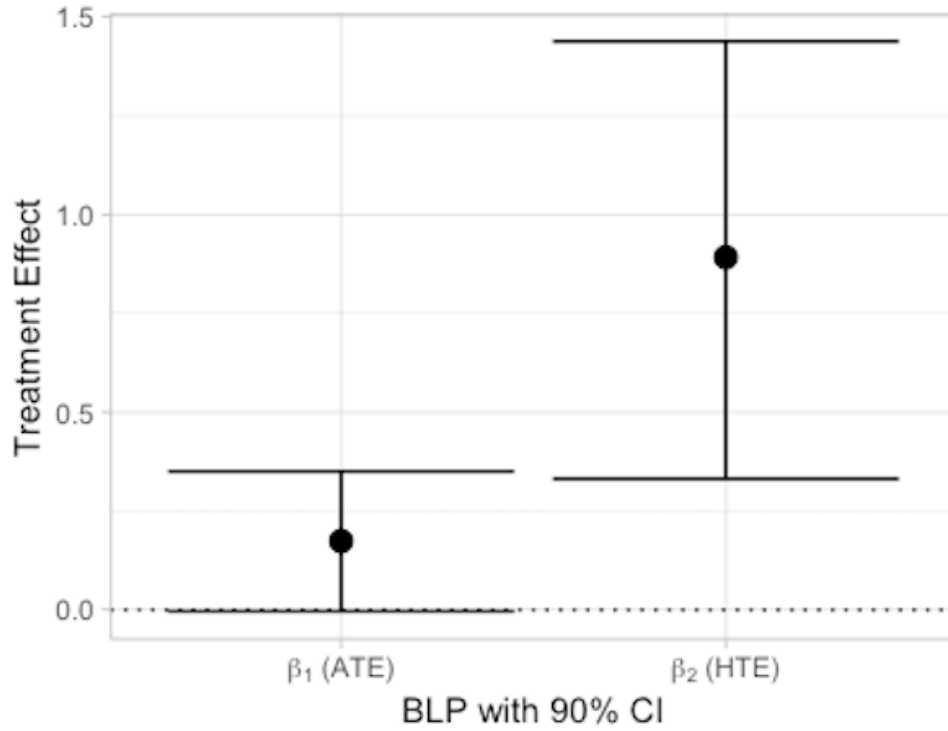
Table A.5.5: Treatment Effects on Endline English Test Scores

	(1) Standardised Scores	(2) Raw Scores
Expectations	-0.183* (0.108)	-2.307 (1.750)
Peer	-0.235** (0.111)	-2.813 (1.720)
Information	-0.070 (0.113)	-0.481 (1.789)
Constant	0.352** (0.151)	84.118*** (2.383)
Exp vs Peer	0.559	0.701
Exp vs Info	0.200	0.175
Info vs Peer	0.066	0.069
Observations	1222	1224
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The scores in column (1) are standardized using the mean and standard deviation of the historical scores in control schools from administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for the student's baseline score. Regressions include strata fixed effects and standard errors are clustered at the level of randomization.

A.6 Heterogeneity Results

Figure A.6.1: Heterogeneous Treatment Effect of the Peer Arm



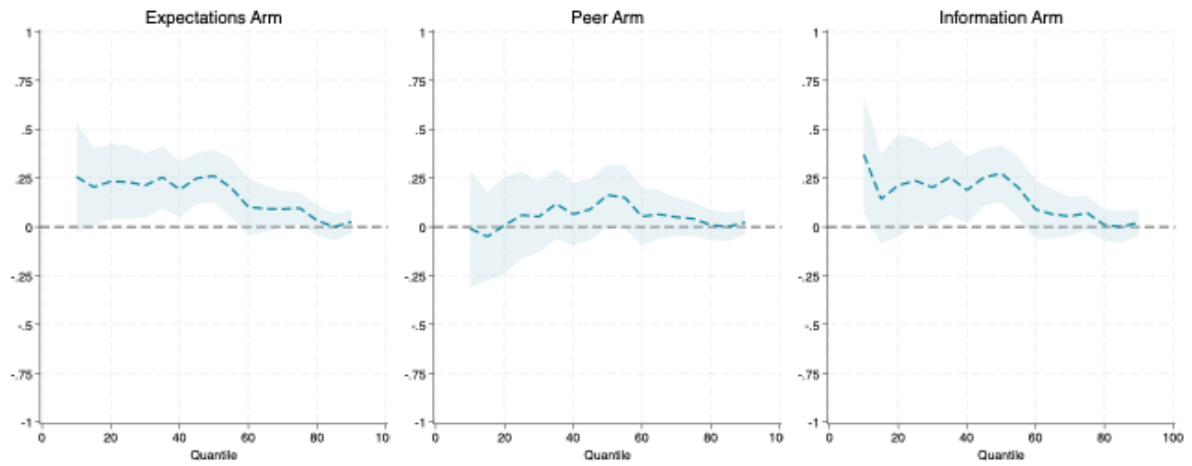
Note: We employ the method in [Chernozhukov et al. \(2018\)](#) and find evidence of heterogeneity in the peer arm as shown by the significantly positive value of β_2 for the best linear predictor. We specify three learners– Lasso, SVM, and Random Forest. The results plotted above correspond to the best linear predictor out of these. Variables for heterogeneity include baseline math performance, gender, whether the matched peer has the same, higher or lower baseline score, and whether the matched peer receives the same, higher, or lower expectation. Standard errors are clustered at the class-level and strata fixed effects are included.

Table A.6.1: Treatment Effects by Baseline Performance of Matched Peer

	(1) Standardised Scores	(2) Percentage Score
Expectations	0.180** (0.073)	3.290** (1.366)
Information	0.148* (0.078)	2.851** (1.431)
Peer score is higher	0.011 (0.104)	-0.025 (1.910)
Peer score is lower	0.057 (0.088)	0.827 (1.602)
Peer score is same	0.173* (0.104)	3.160* (1.738)
Constant	-1.857*** (0.200)	41.602*** (3.815)
Observations	2400	2400
Expectations v/s Peer (Same)	0.941	0.929
Expectations v/s Peer(High)	0.054	0.044
Expectations v/s Peer (Low)	0.097	0.065
Peer (High) vs Peer (Same)	0.180	0.138
Peer (Low) vs Peer (Same)	0.292	0.196
Peer (High) vs Peer (Low)	0.611	0.624
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The scores in column (1) are standardized using historical administrative data by using the mean and standard deviation of the historical scores in control schools. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for the student's baseline score. Regressions control for individuals own score and include strata and round fixed effects. Standard errors are clustered at the level of randomization.

Figure A.6.1: Treatment Effect by Quantiles of Baseline Math Performance.



Note: The figure plots treatment effects on standardised scores for the 10th to 90th quantile in gaps of 5. The shaded area represents the 90% confidence intervals.

Table A.6.2: Heterogeneous Treatment Effects by Predicted Performance:
Leave One Out and Repeated Split Sample Results

(a) Information Arm

Group	RSS	S.E.	LOO	S.E.
Group 1	0.335***	0.133	0.292*	0.174
Group 2	0.190**	0.097	0.201	0.160
Group 3	0.073	0.100	-0.008	0.150
Group 4	-0.081	0.085	0.024	0.106

(b) Expectations Arm

Group	RSS	S.E.	LOO	S.E.
Group 1	0.306***	0.110	0.271**	0.131
Group 2	0.172**	0.086	0.287**	0.138
Group 3	0.215***	0.079	0.105	0.122
Group 4	0.004	0.075	0.049	0.092

(c) Peer Arm

Group	RSS	S.E.	LOO	S.E.
Group 1	0.151	0.115	0.088	0.138
Group 2	0.018	0.088	0.114	0.130
Group 3	0.067	0.078	0.007	0.122
Group 4	-0.046	0.085	-0.011	0.106

Note: This table shows heterogeneous treatment effects for four different groups of students classified based on their predicted math scores. Group 1 includes those predicted to perform the worst and Group 4 is predicted to perform the best. We implement the procedure in [Abadie et al. \(2018\)](#) to predict heterogeneous effects. RSS refers to the coefficient as per the repeated split sample estimator and LOO refers to the coefficient as per the leave-one-out estimator i.e. the effect of the treatment arm on the test scores of students in the particular group. The column labeled S.E. displays the respective standard errors.

Table A.6.3: Heterogenous Treatment Effects by Schools that Share Printed Report Cards

	(1) Standardised Score	(2) Percentage Score
Expectations	0.237** (0.093)	4.651*** (1.651)
Peer	0.111 (0.097)	2.764* (1.633)
Information	0.198* (0.105)	3.839** (1.780)
Printed Report Card	-0.029 (0.114)	3.061 (2.335)
Expectations × Printed Report Card	-0.109 (0.129)	-3.454 (2.702)
Peer × Printed Report Card	-0.111 (0.148)	-4.342 (2.677)
Information × Printed Report Card	-0.075 (0.151)	-2.751 (2.880)
Constant	-0.146 (0.110)	39.159*** (3.580)
Observations	2773.000	2640.000

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The estimations pool midline and endline scores. The scores in column (1) are standardized using the mean and standard deviation of the historical scores in control schools from administrative data. Column (2) reports the raw scores of students (converted to percentages) in a value-added specification i.e. controlling for the student's baseline score. Regressions include strata and round fixed effects. Standard errors are clustered at the level of randomization.

A.7 Robustness and Alternative Specifications

Table A.7.1: Treatment Effects using PDS Lasso

	(1) Standardised Score	(2) Percentage Score
Information Arm	0.137* (0.074)	2.581* (1.473)
Expectations Arm	0.213*** (0.068)	4.030*** (1.322)
Peer Arm	0.074 (0.071)	1.201 (1.292)
Constant	-1.993*** (0.195)	33.469*** (4.086)
Observations	2676	2676
Exp vs Peer	0.007	0.003
Info vs Peer	0.306	0.251
Exp vs Info	0.178	0.223
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered on the classroom level. The estimations pool midline and endline scores. Regressions include strata and round fixed effects. Missing values of any baseline characteristics are imputed to be equal to the mean value of the characteristic in the class.

Table A.7.2: Heterogenous Treatment Effects by Homophily using PDS Lasso

	(1) Standardised Scores	(2) Percentage Score
Information	0.140* (0.074)	2.595* (1.441)
Expectations	0.213*** (0.067)	3.995*** (1.310)
Peer	-0.223 (0.187)	-4.347 (3.728)
Peer x Homophily	0.405* (0.218)	7.580* (4.377)
Constant	-1.952*** (0.197)	37.022*** (4.013)
Observations	2430	2430
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered on the classroom level. The estimations pool midline and endline scores. Regressions include strata and round fixed effects. Missing values of any baseline characteristics are imputed to be equal to the mean value of the characteristic in the class.

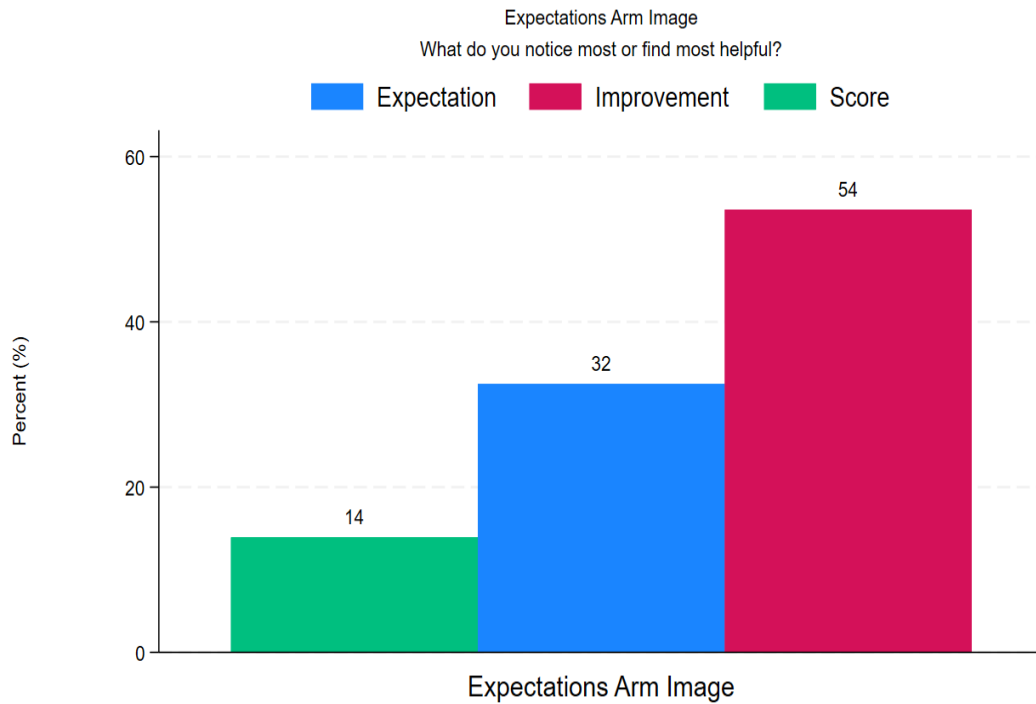
Table A.7.3: Treatment Effects Using an Alternative Specification

	(1) Standardised Score	(2) Percentage Score
Information	0.161** (0.077)	2.747* (1.435)
Expectations (Exp & Peer Arm)	0.186*** (0.066)	3.261** (1.374)
Peer Match	-0.117** (0.059)	-2.175** (0.977)
Observations	2773	2640
Exp vs Peer	0.003	0.006
Exp vs Info	0.688	0.639
Info vs Peer	0.005	0.005
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the classroom level. The estimations pool midline and endline scores. Regressions include strata and round fixed effects. In this specification 'Information' takes a value of 1 for the Information arm, 'Expectations' takes a value of 1 for *both* the Expectations and Peer arm, and 'Peer Match' takes a value of 1 only for the Peer Arm.

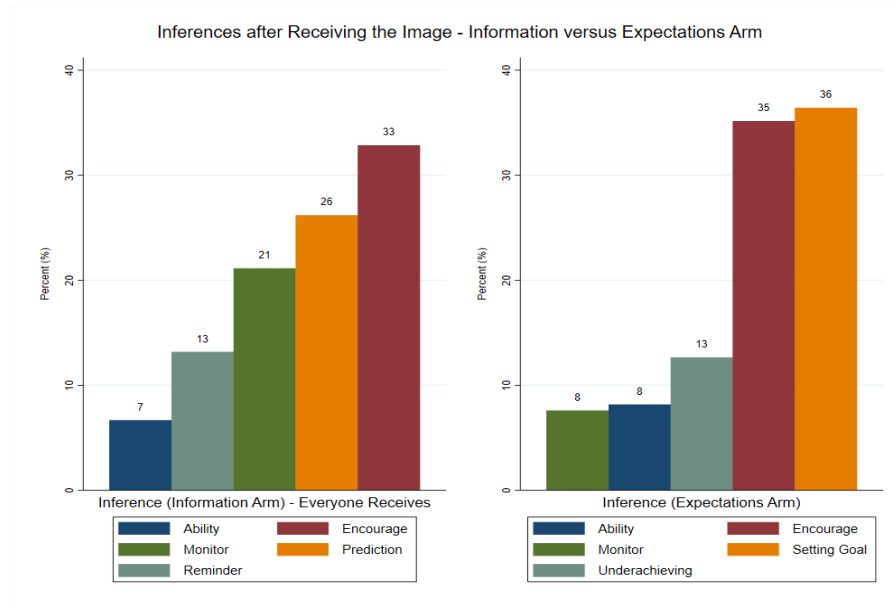
A.8 Follow-up Survey Results

Figure A.8.1: What Students Notice in the Expectations Arm Image

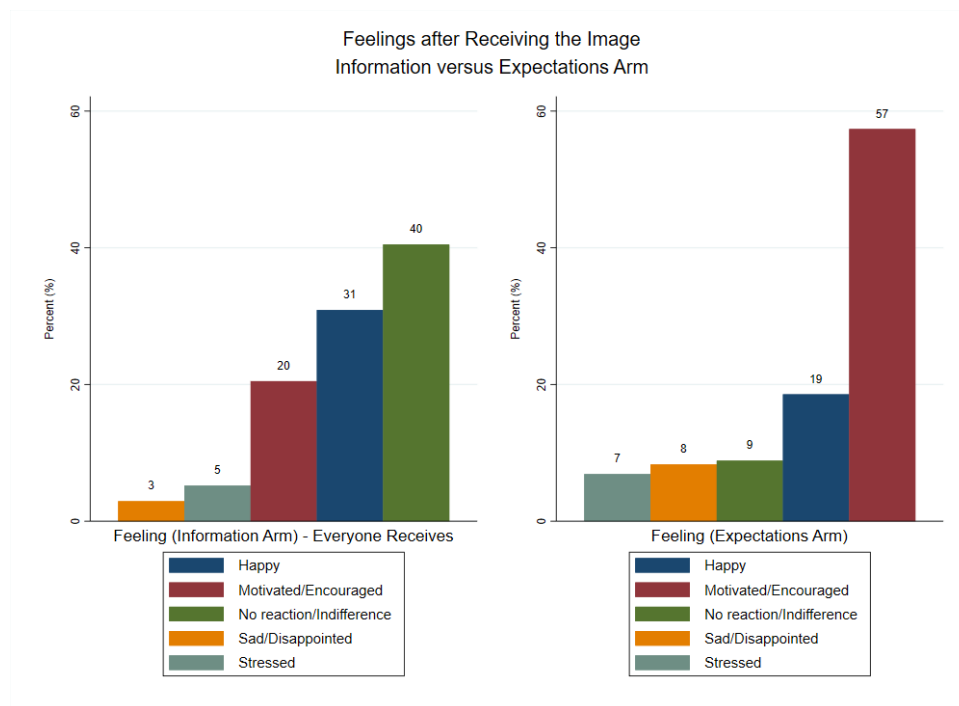


Note: The student follow-up survey sample size was 997 students. The figure illustrates survey responses to the question: 'What do you notice most or find most helpful in this picture?' Respondents had three options: 'Information about your current performance' (labeled as 'Score'), 'How much I can improve and tips on how to get there' (labeled as 'Improvement'), and 'What my teacher thinks I can achieve' (labeled as 'Expectation').

Figure A.8.2: Inferences and Feelings about Images - Expectations and Information Arms



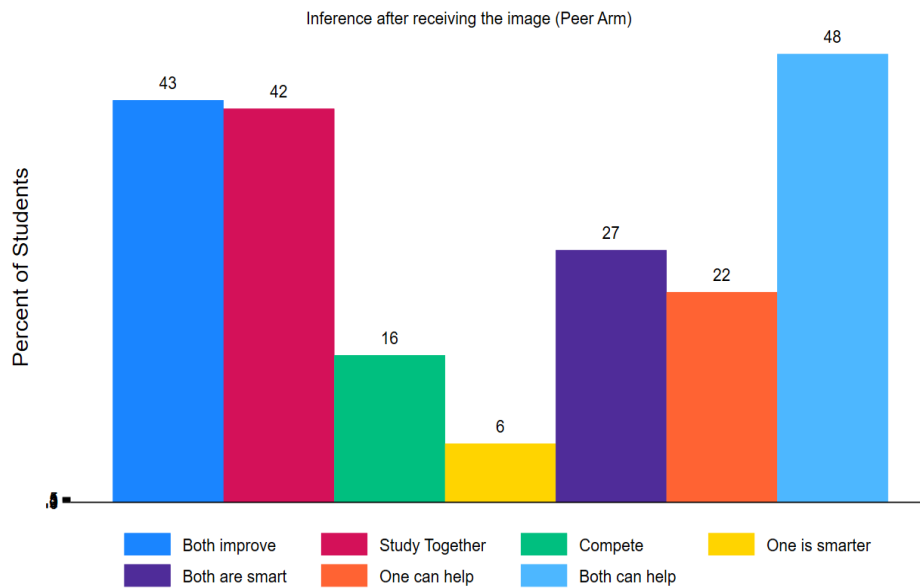
(a) Student Inferences from the Information and Expectations Arm Images



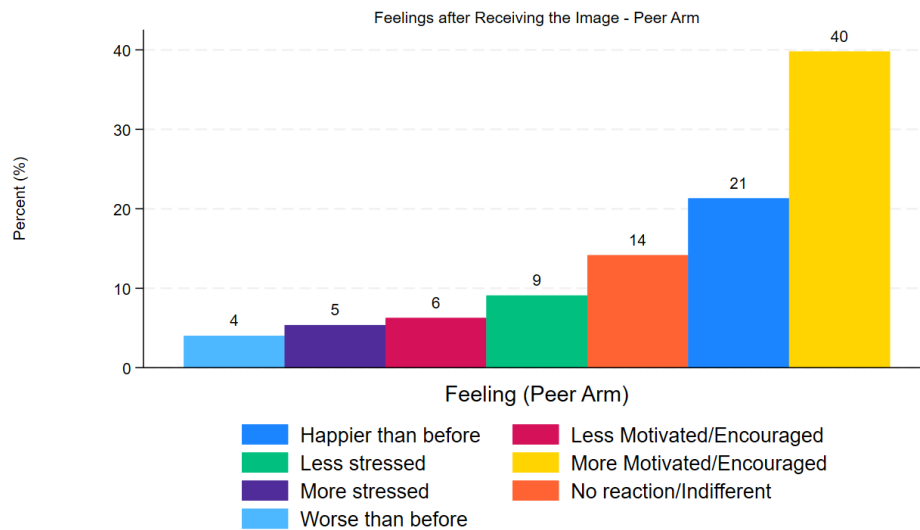
(b) Student Feelings about the Information and Expectations Arm Images

Note: Panel (a) presents students' thoughts after receiving the image in the Information Arm (left) and Expectations Arm (right). Respondents could choose from: 'My teacher is monitoring my progress' (labeled 'monitor'), 'My teacher is encouraging me to do better' (labeled 'encourage'), 'My teacher wants to communicate how smart she thinks I am' (labeled 'ability'), 'My teacher is helping me set a goal to achieve' (labeled 'setting goal'), 'My teacher thinks I am not currently fulfilling my potential' (labeled 'underachieving'), 'My teacher is reminding me of my math score' (labeled 'reminder') and 'My teacher expects me to continue achieving this score' (labeled 'prediction'). Panel (b) figure presents students' reactions when asked how they would feel if they received the image in the Information Arm (left) and Expectations Arm (right). 54

Figure A.8.3: Inferences and Feelings about Images - Peer Arm



(a) What Students Infer from the Peer Arm Image



(b) What Students Feel about the Peer Arm Image

Note: Panel (a) presents students' thoughts after receiving the image in the Peer Arm. Respondents chose from: "My teacher is encouraging us to do better" (labeled as 'Both improve'), "My teacher is encouraging us to work or study together" (labeled as 'Study together'), "My teacher thinks we should compete with each other" (labeled as 'Compete'), "My teacher thinks one of us is smarter than the other" (labeled as 'One is smarter'), "My teacher thinks both of us are equally smart" (labeled as 'Both are smart'), "My teacher thinks one of us can help the other" (labeled as 'One can help'), and "My teacher thinks we both can help each other" (labeled as 'Both can help'). Panel (b) figure presents students' reactions when asked how they would feel if they received the image in the Peer Arm (left) compared to just receiving the Expectations Arm Image.

A.9 Cost-effectiveness Calculation

Table A.9.1: Cost-effectiveness Calculation

	Description	Value
A	Total cost of the design of the infographic images for all treatment arms	\$175
B	Total number of students in treatment arms at endline	1047
C	Design cost per student (A/B)	\$0.17
D	Expectations Arm Treatment Effect (s.d.)	0.186
E	Information Arm Treatment Effect (s.d.)	0.161
F	0.1 s.d. increase cost in the Expectations Arm ($C/D \times 0.10$)	\$0.091
G	0.1 s.d. increase cost in the Information Arm ($C/E \times 0.10$)	\$0.105

Note: The Table calculates the per-student unit cost of a 0.1 standard deviation increase in test scores to aid comparisons with the literature. As we delivered the images in the Expectations, Information, and Peer Arm, the total cost of design (in Row A) is divided by the total number of students in all these three arms (in Row B) to arrive at the per-student cost of designing this info-graphic (Row C).

B Theoretical Appendix

B.1 Model

We present a simple model to show how teacher expectations, information provision, and peer motivation can affect student motivation and achievement.

A student's utility is:

$$U_i = U(s_i, e_i) = s_i - c(e_i) \quad (1)$$

where e_i represents student effort, s_i denotes student test scores, and cost of effort $c(e_i)$ is increasing and convex i.e. $c'_e > 0$ and $c''_e > 0$. For simplicity, we can assume that the cost function $c(e_i) = \frac{1}{2}e^2$ even though the results will hold for any increasing, convex cost function. $U'_s \geq 0$ i.e. utility is rising in test scores.

Student test scores are increasing in student's ability a_i , student motivation m_i , and student effort e_i . For simplicity, we use the following functional form that allows for complementarities between all these components:

$$s_i = a_i \cdot m_i \cdot e_i \quad (2)$$

Finally, student motivation m_i depends on intrinsic motivation \bar{m}_i , and extrinsic motivation, which in turn, depends on teachers and peers. Without loss of generality, we normalize extrinsic motivation to be m_0 for those in our control group.

We do not take an ex-ante stance on whether our treatment arms positively or negatively affect extrinsic motivation. For those in the Information Arm, we assume that extrinsic motivation depends on the interpretation of the receipt of the reminder about the last test score r_i . For those in the Expectations Arm, we assume that the student's extrinsic motivation depends on $h_i - s_{it-1}$ i.e. the gap in the teacher's expectation (h_i) compared to the student's last performance s_{it-1} . Even though this gap is non-negative on average by design in our experiment, it can motivate or frustrate students. Finally, for those in the Peer Arm, extrinsic motivation additionally depends on $h_i - h_{ip}$ i.e. the gap in the teacher's expectation for the student (h_i) compared to their matched peer's (h_{ip}). Since comparisons with the matched peer can motivate or discourage effort depending on how the information is interpreted, it is ambiguous how this affects student motivation.²¹

²¹We have assumed for simplicity of exposition that randomly matched peers only compare their teachers' expectations with each other but more generally, the comparison could also involve other characteristics such as baseline performance and parental wealth, etc. The finding would generalize if we therefore instead considered an index representing the dissimilarity of the matched peers in a student's motivation function instead of just the gap between their expectations. Empirically, our Peer Arm heterogeneity analysis helps shed light on which characteristics matter.)

We can therefore represent the student's motivation function as:

$$m_i = \begin{cases} \bar{m}_i + m_0, & \text{if } i \text{ is in Control Group.} \\ \bar{m}_i + m_0 + f(r_i, h_i - s_{it-1}, h_i - h_{ip}), & \text{if } i \text{ is in Treatment Group} \end{cases} \quad (3)$$

We assume the effect of r_i , $h_i - s_{it-1}$ and $h_i - h_{ip}$ on f is additively separable. In the Information Arm, $h_i - s_{it-1} = 0$ and $h_i - h_{ip} = 0$ as there is no teacher expectation communicated. Similarly, given the experimental design, $h_i - h_{ip} = 0$ and $r_i = 0$ for the Expectations Arm, and $r_i = 0$ for the Peer Arm. We denote the derivatives of f with respect to its three inputs r_i , $h_i - s_{it-1}$, and $h_i - h_{ip}$ by f'_1 , f'_2 , and f'_3 respectively.

This framework yields the following propositions.²²

Proposition 1 (Expectations Arm Effect on Test Scores) *The treatment effect of the Expectations Arm on a student's test scores is positive and increasing in the teacher's expectation, h_i , if and only if student motivation is increasing in teacher expectations i.e. $f'_2 \geq 0$.*

The proposition shows that a comparison of the Expectations Arm to the Control Group empirically can give us an insight into whether high teacher expectations relative to one's performance enter the student motivation function positively.

The proposition below shows that the Peer Arm effect on a student's test score is ambiguous.

Proposition 2 (Peer Arm Effect on Test Scores) *(i) The treatment effect of the Peer Arm on a student's test scores can be positive or negative depending on how teacher expectations of the student compare with that of their matched peer $h_i - h_p$, whether this is motivating ($f'_3 > 0$) or discouraging ($f'_3 < 0$), and whether receiving high teacher expectation is motivating ($f'_2 > 0$) or not ($f'_2 < 0$). (ii) The treatment effect is strictly decreasing in the peer's expectation if and only if receiving a higher teacher expectation than one's peer is motivating, i.e., $f'_3 > 0$.*

Part (i) of the proposition shows that the treatment effect of the Peer Arm is ambiguous ex-ante and depends on the signs of f'_2 , f'_3 and $h_i - h_p$. For example, an intuitive case (and sufficient condition) for the Peer Arm treatment effect to be positive is if the motivation is increasing in teacher expectations i.e. $f'_2 > 0$, in the student's relative comparison of teacher expectations with their peer i.e. $f'_3 > 0$, and if the teacher's expectation for the student is greater than that for their matched peer's i.e. $h_i \geq h_p$. However, if $f'_2 > 0$ and $f'_3 > 0$ as before, but the teacher's expectation for the student is less than their matched peer's, that is, $h_i \leq h_p$, and the resulting discouragement effect is larger than the positive effect of receiving a high expectation from the teacher, the effect of the Peer Arm can be negative on a student's test score.

²²All the proofs are in Appendix B.2.

Therefore, part (i) of the proposition highlights that we expect the treatment effect of the Peer Arm to be heterogeneous across students who have higher or lower teacher expectations relative to their matched peers. Further, the comparison of the treatment effect of the Peer Arm to that of the Expectations Arm can provide insights into the relative strengths of the effect of receiving high teacher expectations (on average) compared to the relative peer comparisons effect empirically. Part (ii) of the proposition illustrates that we can infer whether relative peer comparisons have a positive or discouraging effect on student motivation (on average) by looking at the heterogeneity of the Peer Arm treatment effect by the gap between a student's teacher's expectation and their matched peer's.

Finally, the proposition below shows that the treatment effect of the information arm depends on how the reminder about the last score is interpreted by the student.

Proposition 3 (Information Arm Effect on Test Scores) *The treatment effect of the Information Arm on a student's test scores is positive following a reminder about the last test score if and only if the message is interpreted to be encouraging, that is, $f(r_i, 0, 0) > 0$.*

The proposition shows that a comparison of the Information Arm to the Control Group can empirically help us evaluate whether receiving a reminder about one's last test score from the teacher is perceived to be encouraging and enters the student's motivation function positively.

B.2 Proofs

Proof 1 We drop the i subscript for ease of notation.

From the utility function, we have $U(s, e) = s - c(e)$, where $s = a \cdot m \cdot e$. Let e_0 and e_1 represent the optimal effort levels of a student in the Control Group and the Expectations Arm, respectively. Thus, the treatment effect can be expressed as

$$\Delta s_{Exp} = a \cdot m_1 \cdot e_1 - a \cdot m_0 \cdot e_0$$

To derive the optimal effort level, we maximize the utility function by setting the derivative with respect to effort ($\frac{\partial U}{\partial e}$) equal to zero:

$$\begin{aligned} \frac{\partial U}{\partial e} &= \frac{\partial}{\partial e}(s - c(e)) = 0 \\ \frac{\partial s}{\partial e} - \frac{\partial c(e)}{\partial e} &= 0 \\ a \cdot m - c'(e) &= 0 \\ a \cdot m - e &= 0 \\ e &= a \cdot m \end{aligned}$$

Substituting this optimal effort level into the test scores equation, we get:

$$\begin{aligned} s &= a \cdot m \cdot e \\ &= a \cdot m \cdot (a \cdot m) \\ &= a^2 \cdot m^2 \end{aligned}$$

Therefore, the treatment effect Δs_{Exp} can be expressed as:

$$\begin{aligned} \Delta s_{Exp} &= s_1 - s_0 \\ &= a^2 \cdot m_1^2 - a^2 \cdot m_0^2 \\ &= a^2 \cdot (m_1^2 - m_0^2) \\ &= a^2 \cdot ((\bar{m} + m_o + f(\cdot, h_i - s_{t-1}, \cdot))^2 - (\bar{m} + m_o)^2) \end{aligned}$$

When the motivation function is positive and increasing in the gap between the teacher's expectation

and the student's last performance i.e. $f'_2 > 0$, the treatment effect Δs_{Exp} is positive (and vice versa).

To show that the treatment effect is increasing in h_i , we examine the derivative of the treatment effect with respect to h_i :

$$\frac{\partial \Delta s_{Exp}}{\partial h_i} = \frac{\partial s_1}{\partial h_i} - \frac{\partial s_0}{\partial h_i}$$

Since $\frac{\partial s_0}{\partial h_i} = 0$, differentiating the test scores equation with respect to h_i , we get:

$$\frac{\partial s_1}{\partial h_i} = \frac{\partial}{\partial h_i}(a^2 \cdot m_1^2) = a^2 \frac{\partial}{\partial h_i}(m_1^2) = a^2 \cdot 2 \frac{\partial f}{\partial h_i}$$

Therefore, if $f'_2 > 0$, we find that the treatment effect Δs_{Exp} is increasing in h_i (and vice versa).

Proof 2 Similar to the previous proof, we derive the optimal effort level and treatment effect. Let e_0 and e_2 represent the optimal effort levels of a student in the Control Group and the Peer Arm, respectively. The treatment effect can be expressed as $\Delta s_{Peer} = a \cdot m_2 \cdot e_2 - a \cdot m_0 \cdot e_0$.

As, before, solving for the optimal effort level by maximizing the utility function with respect to e gives $e = a \cdot m$. Therefore, the treatment effect Δs_{Peer} can be expressed as:

$$\begin{aligned} \Delta s_{Peer} &= a^2 \cdot m_2^2 - a^2 \cdot m_0^2 \\ &= a^2 \cdot (m_2^2 - m_0^2) \\ &= a^2 \cdot ((\bar{m} + m_o + f(:, h_i - s_{t-1}, h_i - h_p))^2 - (\bar{m} + m_o)^2) \end{aligned}$$

Consider the case where $f'_2 > 0$ and so the gap between one's teacher's expectation and own performance always has a positive effect. Now, if $f'_3 > 0$ and $h_i \geq h_p$ i.e. motivation is increasing in the gap of one's own teacher expectation relative to the peer and this gap is non-negative, $\Delta s_{Peer} > 0$. If $h_i = h_p$ i.e. the peers have similar teacher expectations, then also we will have $\Delta s_{Peer} > 0$. However, $f'_3 \geq 0$ and $h_i < h_p$, then Δs_{Peer} can be positive or negative. This will depend on the relative strength of the motivational effect from high teacher expectation and the discouragement effect from disparity in teacher expectation with the peer i.e. whether $f'_2 h_i$ is less than or greater than $f'_3 (h_i - h_p)$. If $f'_2 h_i < f'_3 (h_i - h_p)$, then the effect of the peer arm will be negative as the positive effect of teacher expectations is not as strong as the negative effect of peer comparisons.

To analyze the derivative of the treatment effect with respect to h_i and h_p , we differentiate Δs_{Peer} with respect to h_i and h_p separately.

1. Derivative with respect to h_i :

$$\begin{aligned}
\frac{\partial \Delta S_{Peer}}{\partial h_i} &= \frac{\partial}{\partial h_i} (a^2 \cdot (m_2^2 - m_0^2)) \\
&= a^2 \cdot (2m_2 \cdot \frac{\partial m_2}{\partial h_i} - 2m_0 \cdot \frac{\partial m_0}{\partial h_i}) \\
&= a^2 \cdot 2m_2 \cdot \frac{\partial m_2}{\partial h_i} \\
&= a^2 \cdot 2m_2 \cdot (f'_2(h_i - s_{t-1}, h_i - h_p) \frac{\partial(h_i - s_{t-1})}{\partial h_i} + f'_3(h_i - s_{t-1}, h_i - h_p) \frac{\partial(h_i - h_p)}{\partial h_i}) \\
&= a^2 \cdot 2m_2 \cdot (f'_2(h_i - s_{t-1}, h_i - h_p) - f'_3(h_i - s_{t-1}, h_i - h_p))
\end{aligned}$$

2. Derivative with respect to h_p :

$$\begin{aligned}
\frac{\partial \Delta S_{Peer}}{\partial h_p} &= \frac{\partial}{\partial h_p} (a^2 \cdot (m_2^2 - m_0^2)) \\
&= a^2 \cdot (2m_2 \cdot \frac{\partial m_2}{\partial h_p} - 2m_0 \cdot \frac{\partial m_0}{\partial h_p}) \\
&= a^2 \cdot (2m_2 \cdot \frac{\partial m_2}{\partial h_p}) \\
&= a^2 \cdot 2m_2 \cdot f'_3(h_i - s_{t-1}, h_i - h_p) \frac{\partial(h_i - h_p)}{\partial h_p} \\
&= a^2 \cdot 2m_2 \cdot -f'_3(h_i - s_{t-1}, h_i - h_p)
\end{aligned}$$

Since f'_2 , and f'_3 can be positive or negative depending on the specific functional form of f , we cannot draw a definitive conclusion about the signs of $\frac{\partial \Delta S_{Peer}}{\partial h_i}$ and $\frac{\partial \Delta S_{Peer}}{\partial h_p}$. However, if $f'_3 > 0$, then $\frac{\partial \Delta S_{Peer}}{\partial h_p} < 0$ (and vice-versa).

Therefore, the treatment effect ΔS_{Peer} can be positive or negative, depending on the relative strength of the motivational effect from high teacher expectation and the discouragement effect from the disparity in teacher expectation with the peer, as well as the specific functional form of the motivation function and its derivatives with respect to h_i and h_p .

Proof 3 Similar to the previous proofs, we derive the optimal effort level and treatment effect. Let e_0 and e_3 represent the optimal effort levels of a student in the Control Group and the Information Arm, respectively. The treatment effect can be expressed as $\Delta S_{Info} = a \cdot m_3 \cdot e_3 - a \cdot m_0 \cdot e_0$.

As before, solving for the optimal effort level by maximizing the utility function with respect to e gives $e = a \cdot m$. Therefore, the treatment effect ΔS_{Info} can be expressed as:

$$\begin{aligned}
\Delta s_{Info} &= a^2 \cdot m_3^2 - a^2 \cdot m_0^2 \\
&= a^2 \cdot (m_3^2 - m_0^2) \\
&= a^2 \cdot ((\bar{m} + f(r))^2 - \bar{m}^2)
\end{aligned}$$

Thus, if the motivation effect of the reminder is positive (negative), the treatment effect Δs_{Info} is positive (negative).

References

- Abadie, A., Chingos, M. M. and West, M. R. (2018), 'Endogenous stratification in randomized experiments', *Review of Economics and Statistics* **100**(4), 567–580.
- Alan, S., Boneva, T. and Ertac, S. (2019), 'Ever failed, try again, succeed better: Results from a randomized educational intervention on grit', *The Quarterly Journal of Economics* **134**(3), 1121–1162.
- Anderson, M. L. (2008), 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects', *Journal of the American statistical Association* **103**(484), 1481–1495.
- Andrabi, T., Das, J. and Khwaja, A. I. (2017), 'Report cards: The impact of providing school and child test scores on educational markets', *American Economic Review* **107**(6), 1535–63.
- Andrabi, T., Das, J., Khwaja, A. I., Vishwanath, T. and Zajonc, T. (2007), 'Learning and educational achievements in punjab schools (leaps): Insights to inform the education policy debate', *World Bank, Washington, DC*.
- Angrist, J. D., Pathak, P. A. and Walters, C. R. (2013), 'Explaining charter school effectiveness', *American Economic Journal: Applied Economics* **5**(4), 1–27.
URL: <https://www.aeaweb.org/articles?id=10.1257/app.5.4.1>
- Barrera-Osorio, F., Gonzalez, K., Lagos, F. and Deming, D. J. (2020), 'Providing performance information in education: An experimental evaluation in colombia', *Journal of Public Economics* **186**, 104185.
- Bau, N. and Das, J. (2020), 'Teacher value added in a low-income country', *American Economic Journal: Economic Policy* **12**(1), 62–96.
- Belloni, A., Chernozhukov, V. and Hansen, C. (2014), 'Inference on treatment effects after selection among high-dimensional controls', *Review of Economic Studies* **81**(2), 608–650.
- Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006), 'Adaptive linear step-up procedures that control the false discovery rate', *Biometrika* **93**(3), 491–507.
- Berry, J. (2015), 'Child control in education decisions: An evaluation of targeted incentives to learn in india', *Journal of Human Resources* **50**(4), 1051–1080.
- Beteille, T. and Evans, D. K. (2019), *Successful teachers, successful students: Recruiting and supporting society's most crucial profession*, World Bank Group Washington, DC.
- Blimpo, M. P. (2014), 'Team incentives for education in developing countries: A randomized field experiment in benin', *American Economic Journal: Applied Economics* **6**(4), 90–109.

- Burke, M. A. and Sass, T. R. (2013), 'Classroom peer effects and student achievement', *Journal of Labor Economics* **31**(1), 51–82.
- Bursztyn, L., Egorov, G. and Jensen, R. (2019), 'Cool to be smart or smart to be cool? understanding peer pressure in education', *The Review of Economic Studies* **86**(4), 1487–1526.
- Bursztyn, L., Fujiwara, T. and Pallais, A. (2017), "acting wife": Marriage market incentives and labor market investments', *American Economic Review* **107**(11), 3288–3319.
- Bursztyn, L. and Jensen, R. (2015), 'How does peer pressure affect educational investments?', *The quarterly journal of economics* **130**(3), 1329–1367.
- Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2009), 'Peer effects and social networks in education', *The review of economic studies* **76**(4), 1239–1267.
- Carlana, M. (2019), 'Implicit stereotypes: Evidence from teachers' gender bias', *The Quarterly Journal of Economics* **134**(3), 1163–1224.
- Chernozhukov, V., Demirer, M., Duflo, E. and Fernandez-Val, I. (2018), Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india, Technical report, National Bureau of Economic Research.
- Chetty, R., Friedman, J. N. and Rockoff, J. E. (2014a), 'Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates', *American economic review* **104**(9), 2593–2632.
- Chetty, R., Friedman, J. N. and Rockoff, J. E. (2014b), 'Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood', *American economic review* **104**(9), 2633–2679.
- Crouch, L., Kaffenberger, M. and Savage, L. (2021), 'Using learning profiles to inform education priorities: An editors' overview of the special issue', *International Journal of Educational Development* **86**, 102477.
- Damgaard, M. T. and Nielsen, H. S. (2018), 'Nudging in education', *Economics of Education Review* **64**, 313–342.
- De Ree, J., Muralidharan, K., Pradhan, M. and Rogers, H. (2018), 'Double for nothing? experimental evidence on an unconditional teacher salary increase in indonesia', *The Quarterly Journal of Economics* **133**(2), 993–1039.
- Djaker, S., Ganimian, A. J. and Sabarwal, S. (2022), 'Out of sight, out of mind? the gap between students' test performance and teachers' estimations in india and bangladesh', *Unpublished manuscript*.

- Dobronyi, C. R., Oreopoulos, P. and Petronijevic, U. (2019), 'Goal setting, academic reminders, and college success: A large scale field experiment', *Journal of Research on Educational Effectiveness* **12**(1), 38–66.
- Dozza, L. and Cavrini, G. (2012), 'Perceptions of competence: how parents view teachers', *Procedia-Social and Behavioral Sciences* **46**, 4050–4055.
- Duckworth, A. L. and Quinn, P. D. (2009), 'Development and validation of the short grit scale (grit-s)', *Journal of personality assessment* **91**(2), 166–174.
- Evans, D. K. and Popova, A. (2016), 'What really works to improve learning in developing countries? an analysis of divergent findings in systematic reviews', *The World Bank Research Observer* **31**(2), 242–270.
- Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K. and Trautwein, U. (2015), 'Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement', *Contemporary Educational Psychology* **41**, 1–12.
- Fryer Jr, R. G. (2014), 'Injecting charter school best practices into traditional public schools: Evidence from field experiments', *The Quarterly Journal of Economics* **129**(3), 1355–1407.
- Ganimian, A. J. (2020), 'Growth-mindset interventions at scale: Experimental evidence from argentina', *Educational Evaluation and Policy Analysis* **42**(3), 417–438.
- Glewwe, P. and Muralidharan, K. (2016), Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications, in 'Handbook of the Economics of Education', Vol. 5, Elsevier, pp. 653–743.
- Hirshleifer, S. (2015), 'Incentives for effort or outputs? a field experiment to improve student performance'. Working Paper.
- Islam, A., Kwon, S., Masood, E., Prakash, N., Sabarwal, S. and Saraswat, D. (2021), 'When goal-setting forges ahead but stops short: Evidence from an at scale rct in tanzania'.
- Jackson, M. O., Nei, S. M., Snowberg, E. and Yariv, L. (2023), The dynamics of networks and homophily, Technical report, National Bureau of Economic Research.
- Jussim, L. and Harber, K. D. (2005), 'Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies', *Personality and social psychology review* **9**(2), 131–155.
- Kremer, M., Miguel, E. and Thornton, R. (2009), 'Incentives to learn', *The Review of Economics and Statistics* **91**(3), 437–456.

- Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C. and Rajani, R. (2019), 'Inputs, incentives, and complementarities in education: Experimental evidence from tanzania', *The Quarterly Journal of Economics* **134**(3), 1627–1673.
- Morisano, D., Hirsh, J. B., Peterson, J. B., Pihl, R. O. and Shore, B. M. (2010), 'Setting, elaborating, and reflecting on personal goals improves academic performance.', *Journal of applied psychology* **95**(2), 255.
- Muralidharan, K. and Sundararaman, V. (2011), 'Teacher performance pay: Experimental evidence from india', *Journal of political Economy* **119**(1), 39–77.
- Oreopoulos, P. and Petronijevic, U. (2019), The remarkable unresponsiveness of college students to nudging and what we can learn from it, Technical report, National Bureau of Economic Research.
- Rosenthal, R. and Jacobson, L. (1968), 'Pygmalion in the classroom', *The urban review* **3**(1), 16–20.
- Rosenthal, R. and Jacobson, L. (1992), 'Pygmalion in the classroom. expanded edition', *New York: Irvington* p. 382.
- Schippers, M. C., Scheepers, A. W. and Peterson, J. B. (2015), 'A scalable goal-setting intervention closes both the gender and ethnic minority achievement gap', *Palgrave Communications* **1**(1), 1–12.
- UNESCO (2014), 'Efa global monitoring report. teaching and learning: Achieving quality for all'.
- UNESCO (2020), 'Global education monitoring report 2020: Inclusion and education: All means all', 92310038 .
- Wang, S., Rubie-Davies, C. M. and Meissel, K. (2018), 'A systematic review of the teacher expectation literature over the past 30 years', *Educational Research and Evaluation* **24**(3-5), 124–179.
- Wentzel, K. R. (2010), Students' relationships with teachers, in 'Handbook of research on schools, schooling and human development', Routledge, pp. 75–91.
- Wolf, S. and Brown, A. (2023), 'Teacher beliefs and student learning', *Human Development* **67**(1), 37–54.
- World-Bank (2017), *World development report 2018: Learning to realize education's promise*, The World Bank.

- Wu, J., Zhang, J. and Wang, C. (2023), 'Student performance, peer effects, and friend networks: Evidence from a randomized peer intervention', *American Economic Journal: Economic Policy* **15**(1), 510–542.
- Yeager, D. S. and Dweck, C. S. (2012), 'Mindsets that promote resilience: When students believe that personal characteristics can be developed', *Educational psychologist* **47**(4), 302–314.