Improving Skill Production with Peer-Driven Knowledge Sharing in K-12 Schools

Anuj Kumar
University of Florida
Warrington College of Business
Gainesville, FL
akumar1@ufl.edu

Anurag Garg
University of Kansas
School of Business
Lawrence, KA
anurag@ku.edu

February 28, 2025

Abstract

We present experimental evidence on the impact of peer-driven knowledge sharing in enhancing skill development within K-12 classrooms. A technology platform was developed to identify students' knowledge gaps and organize team contests to induce knowledge sharing among team members. We conducted a large-scale field experiment on 3110 students in 86 Grade 3-6 math classes spanning six cities across three Indian states over two academic years. The experiment randomly assigned classes into three groups: team-based classes, where students collaborated in teams on the platform; individual-based classes, where students worked alone; and control classes, with no access to the platform. Our intent-to-treat (ITT) estimates reveal that students in team-based classes achieved cognitive scores 0.25σ and 0.17σ higher than those in control and individual classes, respectively. While our intervention reduced skill dispersion within team-based classes; it improved the performance of students of all abilities – high, medium, and low – indicating that high-ability students benefit from collaborating with low-ability teammates. Our instrumental variable (IV) estimates suggest that feedback on knowledge gaps and subsequent knowledge sharing with team members were key drivers of the academic improvements observed in team-based classes. Competing in teams could also shape students' social preferences for classmates. Team students reduced their generosity, reciprocity, and trust for students in other teams but not for their teammates. The intervention is highly scalable, requiring minimal adjustments to the existing school setup and limited financial investment. This study demonstrates that well-designed peer interactions through technology platforms can lead to higher and more equitable skill development in K-12 classrooms.

Keywords: Skill production, Peer effects, K-12 education, Knowledge diffusion, EdTech, Field experiment.

A randomized controlled trial, registered as ASPredicted # 97643, is available at https://aspredicted.org/YK1_FNG

1 Introduction

Skill development among K-12 students varies due to differences in their learning motivations, abilities, and needs. Educational technology (EdTech) has the potential to improve learning productivity - the rate at which efforts are transformed into academic success - for students with diverse learning requirements (Escueta et al., 2020; Bulman and Fairlie, 2016). A key example of this is personalized computer-aided learning, which assesses students' learning needs and delivers tailored educational content to support their skill development. However, its effectiveness is limited to students who are motivated to engage with these tools and have sufficient support to use them effectively. Many students from lower-income families tend to be less engaged with education, often perform below grade level, and lack adequate educational support. For these students, personalized EdTech alone may not be sufficient without the right incentives and external support. Consequently, past EdTech interventions have incorporated teacher-led personalized instructions and support within structured classes, either during or outside of school hours (Muralidharan, Singh and Ganimian, 2019; Banerjee et al., 2007). Nevertheless, scaling such interventions remains challenging due to limitations in available teaching staff and other school resources.

Besides teachers and school resources, classroom peers are a readily available resource for students. Can peers effectively motivate and assist students with their learning needs in K-12 classrooms? The existing literature in economics presents mixed findings on the impact of mixing students of different abilities on skill development in K-12 schools (Sacerdote, 2011; Epple and Romano, 2011). However, simply grouping peers together may not guarantee productive interactions, which could explain these inconsistent results. For instance, a study at the US Air Force Academy found that low-ability cadets performed worse when mixed with high-ability peers, as they tended to interact more with their peers of similar ability rather than engaging with the higher-ability students (Carrell, Sacerdote and West, 2013). Given these challenges, leveraging peers as an alternative support system is a promising yet underexplored area of research.

In this research, we developed an integrated EdTech platform (EPInc) designed to encourage classroom peers to support each other's learning needs in K-12 classrooms at scale. EPInc incorporates three main features. First, it *identifies students' knowledge gaps* based on their homework performance and offers them personalized content aligned with their specific learning needs. Sec-

ond, it *organizes students into peer groups* by forming balanced teams within classrooms, ensuring that students of varying abilities work together. This allows higher-ability students to assist their teammates in overcoming knowledge gaps. Lastly, the platform fosters *team identities and organizes team contests* within classrooms, using the desire to win as a team to motivate team members to help each other improve their skills.

Our primary objective is to estimate the impact of peer-supported personalized learning through the EPInc platform on skill development in K-12 classrooms. An alternative approach to encouraging peer assistance is to directly motivate students to engage in self-learning using personalized learning technologies. We hypothesize that structured peer-driven learning will lead to greater engagement and skill improvements than self-learning alone.

We conducted a large-scale field experiment across multiple Indian K-12 schools, spanning a wide geographical area over two academic years. The study involved 3,110 students from 86 Grade 3-6 math classes across 13 schools in six cities across three Indian states. At the beginning of the academic year, we administered standardized math tests to assess students' baseline math skills. Additionally, we measured students' non-cognitive skills, including generosity, reciprocity, and trust, through social preference experiments. We randomly assigned classes within each school into three groups: team, individual, and control classes. To form teams, we randomly selected one student from each quartile of the baseline math score distribution of the class, ensuring balanced and exogenous team compositions. To foster a strong team identity, students were given the opportunity to choose their team mascots and names, which were prominently displayed on their EPInc platform home screens.

Teachers provide class instructions per the existing curriculum and pedagogy in the schools in all classes. After a few lectures, they generate base homework on the EPInc platform for team and individual classes. On completion of the base homework, the team students obtain team-wise feedback showing the knowledge gaps of all the team members, but the students receive feedback on their knowledge gaps alone in individual classes. Teachers encourage students to help their teammates so that their team can win tournaments in team classes and to self-study to win tournaments themselves in individual classes. After the base homework, students in both classes receive personalized homework on the EPInc platform that offers them more questions in their knowledge gaps. Control students received an equal number of paper-based homework in school,

but did not receive feedback about their knowledge gaps or more questions on these topics.

After a few homework assignments, teachers organize class tournaments on the EPInc platform for both team and individual classes. In team classes, students earn team ranks based on the overall scores of their team, while individual students receive rankings based on their personal scores in the tournaments. Teachers provide social recognition to tournament winners. This structure incentivizes students in team classes to share knowledge with their teammates to improve their chances of winning. In contrast, students in individual classes are motivated to focus on self-learning to perform well in tournaments. Control group students do not participate in tournaments and instead complete paper-based homework as a substitute for the tournament activities.

A student in our study was offered more than 400 math questions in approximately 16 homework assignments and four tournaments in an academic year. All academic activities were conducted in classrooms during school hours. Control students were exposed to a similar number of academic activities/hours as team and individual students in schools. To identify the effect of peer help (self-help) on skill development, teachers were instructed not to provide personalized instructions or help to students in team (individual) classes.

At the end of the academic year, endline cognitive and non-cognitive assessments, similar to baseline assessments, were performed.[1] We administered endline tests in a paper format to ensure that control students, without access to the EPInc platform, were not at a disadvantage compared to team and individual students.

The baseline cognitive and non-cognitive scores of students (and those within each ability tercile) across the three assignment groups were statistically similar, suggesting that students had comparable math skills and social preferences at the start of the experiment. However, our intent-to-treat estimates, based on class- and student-level data, revealed higher average endline scores for team students compared to control students, while individual students' scores were statistically similar to those of the control group. Specifically, team classes achieved an average endline score that was 0.23σ higher, while individual classes had scores that were not significantly different from those of control classes in our class-level analysis. The dispersion in endline scores in team classes was also lower than in control classes. At the student level, our analysis showed that team students

[1]Baseline and Endline cognitive tests were prepared by a group of outside educators as per grade-level math syllabus of National Council of Educational Research and Training (NCERT) in India

had endline scores 0.25σ higher than control students, and 0.17σ higher than individual students. In all ability terciles, team students scored higher than corresponding control students. We found statistically similar endline scores between individual and control students across all ability terciles.

The EPInc platform was designed to enhance skill development through three key mechanisms: (1) *team feedback* – providing team members with insights into each other's knowledge gaps, (2) *induced knowledge sharing* – motivating team members to assist each other in order to win tournaments, and (3) *improved learning motivation* – leveraging social recognition from tournament success to encourage students to invest in learning to contribute to their team success. We provide suggestive empirical evidence supporting these mechanisms.

While individual students receive feedback on their knowledge gaps and are motivated to self-learn those topics, team students receive feedback on the strengths and weaknesses of all their team members. This additional information enables team students to identify appropriate teammates to seek help from or offer assistance to. We measure self-feedback as the proportion of homework completed by an individual student, and team feedback as the average number of team members who attended homework with a team student. Our data show considerable variation in homework attendance across teams, allowing us to identify the effect of team feedback on skill development. Our instrumental variable (IV) estimates indicate a 0.11σ improvement in team students' endline scores with each unit increase in team feedback. However, we found no significant impact of self-feedback on individual students' skill development.

Team feedback is effective when team members act on it by sharing knowledge. Students in team classes can report which teammate assisted them with homework on the EPInc platform. To assess the impact of these interactions on skill development, we collected data on the total number of peer engagement for each student. Our IV estimates indicate that an additional team member interaction in team classes leads to a 0.25σ improvement in endline scores.

Effective peer assistance in team-based learning environments is contingent upon a team member's proficiency in a given topic. To examine this relationship, we analyze topic-specific homework scores for students in individual and team classes. Our findings indicate that students in team settings achieve higher endline scores in topics where their teammates demonstrated strong performance on homework. This provides further empirical evidence that peer support is a key mechanism driving the enhanced learning outcomes observed in team-based classrooms.

Additionally, winning tournaments can serve as a powerful motivator for students to invest in their education. Typically, the top three winners are recognized on the social leaderboard in classrooms. In individual classes, this leaderboard highlights the three top-ranked students, usually those with high cognitive abilities. In contrast, in team classes, the three top-ranked teams represent a larger number of more diverse students, including students from varying ability levels .[2] This wider recognition of achievements in team classes may motivate more students to share knowledge and increase their learning efforts. Consistent with this idea, our IV estimates suggest that team students with high tournament ranks have higher endline scores than both individual and control students, while those with low ranks performed at statistically similar levels.

Finally, we examine how team-based incentives influence students' social preferences for others. If students identify with their teams, they may enhance their cooperation with teammates and competition with students in other teams (Tajfel et al., 1971; Choi and Bowles, 2007; Bowles and Gintis, 2011). Our findings indicate that team students decrease their social preferences for students in other teams, but their preferences for teammates remain statistically unchanged.

While the impact of technology – especially personalized learning technologies -— on addressing the constraints of human capital development has been well-documented, these technological aids alone often fail to benefit the underprivileged students who need them the most. In developing countries, limited teaching and family resources severely constrain the effectiveness of these technologies. We provide empirical evidence that fostering peer-led knowledge sharing through group contests on a technology platform can lead to significantly higher (0.25σ) and more equitable skill development in K-12 classrooms at scale. Furthermore, this study contributes to the literature by introducing a novel intervention that not only incentivizes but also quantifies productive peer interactions beyond the conventional method of exogenously mixing peers within classrooms, offering a more comprehensive framework for assessing the impact of peer collaboration on learning outcomes.

Our findings highlight that increased feedback and interactions within teams are key drivers of enhanced skill production. By emphasizing team identity -— through displays of team markers like mascots and names – and recognizing team success on EdTech platforms, we can foster greater team cohesion and improve knowledge sharing among team members, leading to higher skill development

[2]Note that winning teams consist of one student from each quartile of the class's baseline math score distribution.

in classrooms.

We also found that encouraging students to self-learn from personalized feedback and content is not effective. This underscores the limitations of individual-focused technological aids, particularly for students from disadvantaged backgrounds who require external support to learn. Our results suggest that integrating other entities, such as families, teachers, peers, educators, and mentors, into technology platforms could better address students' learning needs.

Our intervention requires minimal financial investment and only little changes in school operations, as it leverages current educational resources, curricula, and pedagogical approaches. This makes it highly scalable and suitable for improving skill development in K-12 classrooms, particularly in under-resourced educational systems. Furthermore, our intervention can be expanded to integrate other key players in the K-12 educational ecosystem, such as parents, teachers, mentors, and educators, alongside peers, further enhancing skill production.

The rest of the paper is organized as follows. Section 2 discusses the literature on peer effects and how our platform incentivizes peer interactions from the economic theory of group conflicts and group contests. Section 3 describes our field setup and experimental design. Section 4 describes our data and balance checks. Section 5 outlines the empirical model and main results. Section 5.5 discusses the underlying mechanisms for our results, and Section 6 concludes by discussing the implications of the current work.

2 Peer Effects in Education

Educational technology (EdTech) can identify students' knowledge gaps based on their demonstrated performance and accordingly serve them the needed educational content. For example, personalized EdTech applications can help students fill their knowledge gaps by offering more educational content (lectures and homework) on those topics. However, these applications only help students who are willing to consume the customized content and can either self-learn or have family and school support to learn from it. A vast majority of students are not adequately engaged with education and thus, they may not put in the requisite effort to learn from personalized content. Even if students put in the effort, many of them may be unable to self-learn from such content. For these reasons, most successful personalized learning interventions were given during after-school

hours with extra support from teachers/ educators (Muralidharan, Singh and Ganimian, 2019; Banerjee et al., 2007). However, implementing these solutions at scale is operationally costly and complex, especially in developing countries with scarce educational resources.

Besides teachers and family resources, peers are an abundantly available yet underutilized resource in the K-12 educational ecosystem. Since students spend a significant amount of time with their peers in schools, peers can motivate students to study and provide them with the needed support to improve their skill production in K-12 classrooms.

There exists extensive literature on peer effects in education – please see the detailed review of these studies in Epple and Romano (2011), Sacerdote (2011), and Bishop (2006). Peer effects are broadly defined as the effect of peer background, behavior, and educational outcomes on students' educational outcomes (and behavior). There are two ways in which peers affect students' academic outcomes.

1. Peers' behavior/ability/background can generate positive or negative externality for students. Peers can generate positive externality by directly helping students through collaboration or indirectly helping by asking good questions from teachers, allowing teachers to teach at a brisk pace/higher level, or allowing teachers to give more one-to-one time to students. Peers can also cause negative externality by disrupting the class, forcing teachers to teach at a slower pace/lower level, or wasting teachers' time maintaining discipline.

2. Peers could affect students' learning efforts and outcomes by influencing social norms. Here also, peers can positively or negatively affect students' outcomes. For example, students may study when their peers exhibit better work ethics and discipline in class. In contrast, peers could also dissuade students from investing in education by branding hard-working students as nerdy or "uncool."

Studies in K-12 schools show that the average quality of peers positively affects students' achievements - called linear-in-means peer effects (Sacerdote, 2011). However, the peer effects may vary with the abilities of peers on students of different abilities. Students gain the most by being surrounded by peers of similar abilities (Sacerdote, 2011). Specifically, high-ability students gain from high-ability peers (Hoxby and Weingarth, 2005; Lavy, Silva and Weinhardt, 2009; Imberman, Kugler and Sacerdote, 2012), but their performance drops in classes when mixed with low-ability peers

(Betts and Zau, 2004; Imberman, Kugler and Sacerdote, 2012). These findings have led to the educational policy of tracking (streaming) students in classrooms based on their abilities. While Duflo, Dupas and Kremer (2011) show the benefits of tracking, some studies find no effect of tracking, and others found mixing different ability students in a class beneficial (Betts and Shkolnik, 2000; Lefgren, 2004). Burke and Sass (2013) found that instead of tracking the high-ability students in a class, spreading them across can maximize the total educational production. They discovered that low-ability students benefit the most from high-ability peers. Similarly, Vigdor and Nechyba (2007) found that classroom heterogeneity raises the overall class scores. Thus, the impact of mixing students of different abilities in a class on overall educational production is unclear, and it could vary based on how peers of different abilities interact.

Studies showing the benefit of tracking policy attribute it to the fact that teachers can efficiently teach at the ability levels in classes with students of homogeneous abilities (Duflo, Dupas and Kremer, 2011). Teaching efficiency may certainly suffer with a high level of heterogeneity in students' abilities in classrooms - imagine teaching a class where many students have below-grade-level understanding. However, most classes have significant heterogeneity in students' abilities even after tracking. Thus, understanding how to harness peer effects across students of heterogeneous abilities is crucial.

2.1 Peer Effects in Cognitive and Non-cognitive Skill Production

A significant body of research shows that students' future economic and social well-being is determined by their cognitive and non-cognitive skills (Heckman, 2008, 2007). Therefore, it is important to understand the effect of peers' cognitive and non-cognitive skills on students' overall skill production.

Besides cognitive achievements, peers' personality traits (non-cognitive skills) could also affect students' cognitive achievements. Although cognitive and non-cognitive skills for students may be highly correlated, studies have shown that, after controlling for peers' cognitive skills, their non-cognitive skills still influence students' academic achievements. For example, Golsteyn, Non and Zölitz (2021) show that students perform better in the presence of persistent peers. Shure (2021) examined the effect of peers' "Big Five" personality traits on students' cognitive achievements and found a positive effect of peers' conscientiousness on students' math and language scores. These

findings suggest that mixing high-ability students – presumably with higher conscientiousness and better work ethics – could result in higher average cognitive achievements in classrooms.

Past research also shows that peers' personality traits can also affect students' non-cognitive (character) skills. A large body of research documents strong peer effects on students' non-cognitive behavior, such as drug use, criminal activities, alcohol drinking, smoking, churchgoing, teen pregnancy, and school dropouts. See Gaviria and Raphael (2001) and Sacerdote (2011) for a review of these studies. Xu, Zhang and Zhou (2022) show that a higher proportion of repeating students affect non-repeating students' cognitive achievements– due to the spillover of poor work ethics from repeating to non-repeating students. However, little guidance is available on how peers' desirable non-cognitive skills (such as conscientiousness and extraversion) affect students' corresponding skills.

To summarize, a significant body of research posits that (1) peers' cognitive and non-cognitive skills influence students' cognitive achievement and (2) their non-cognitive skills affect students' non-cognitive skills (behavior). This literature provides evidence for both - negative peer externalities from low-ability and disruptive students and positive peer externalities from high-ability and disciplined ones. How do we design interventions that minimize the negative externality of low-ability (disruptive) peers and maximize the positive externality of high-ability peers in classrooms?

Most studies on peer effects in education neither design nor observe the peer interactions (or peer groups) inside classrooms. A large number of such studies rely on random assignment of peers of different abilities in dorm rooms/classrooms/schools to identify the peer effects (Carrell, Fullerton and West, 2009; Sacerdote, 2001; Imberman, Kugler and Sacerdote, 2012; Stinebrickner and Stinebrickner, 2006).[3] However, such exogenous mixing of students alone may not result in desirable peer interactions if students form their own peer groups endogenously. A study on the United States Air Force Academy, Carrell, Sacerdote and West (2013) shows that assigning lower-ability freshmen to cohorts (squadrons) to optimize their academic performance resulted in the opposite. The low-ability freshmen endogenously formed their homogeneous group within the cohort instead of interacting with their high-ability peers as intended in the policy. Thus, instead of relying on exogenous peer group manipulation, can we incentivize desirable interactions among

[3]Please see Epple and Romano (2011) and Sacerdote (2011) for details of other empirical studies identifying peer effects

students of different abilities in classrooms? The economic theory on group contests and conflicts provides guidance (Konrad, 2009; Kolmar et al., 2013).

2.2 Incentivizing Peer Interactions with Group Contests

Consider a scenario where students of varying abilities are grouped into balanced teams, and educational contests are organized within classrooms between such teams. The team with the highest aggregate score (the sum of individual member scores) is declared the winner of the contest.[4] The probability of a team winning depends on its aggregate performance, which is a function of both its members' abilities and their efforts. Given that the teams are balanced in terms of initial abilities, the outcome of the contests largely hinges on the efforts of team members to assist each other in learning.

The economic theory of group conflicts provides insight into students' behavior in such contests (Kolmar et al., 2013). n recent years, evolutionary economists have presented compelling arguments, supported by empirical evidence, for the evolution of non-kin altruism —- specifically a form of other-regarding preferences —- in group conflict scenarios. This non-kin altruism, termed *parochial altruism*, contrasts with the *homo economicus* model from traditional economic theory. Under parochial altruism, individuals cooperate with in-group members (even when it is costly) and adopt neutral or hostile attitudes toward members of other groups (Choi and Bowles, 2007; Bowles and Gintis, 2011).

The concept of parochial altruism aligns with the social identity theory, which posits that individuals identify with groups based on arbitrary social markers, act cooperatively with in-group members, and hold neutral or hostile views for out-group members (Tajfel et al., 1979). In fact, Tajfel et al. (1971) shows that in-group cooperation cannot be elicited without an out-group. In an economic model of parochial altruism in group conflicts, Wagener and Kolmar (2011) show that, in equilibrium, individuals expend higher efforts under parochial altruism than in the self-interest benchmark, and out-group hostility enhances this overbidding. Experimental studies on group contests provide further evidence for higher individual efforts in group contests than predicted by models based on self-interest axiom (Sheremeta, 2018; Dechenaux, Kovenock and Sheremeta, 2015).

[4] This setup is akin to an all-pay auction contest, where the group with the highest performance wins with certainty (Konrad, 2009).

These findings suggest that, if team markers are thoughtfully implemented in classrooms, students may identify more strongly with their team and invest greater efforts to achieve team objectives.

Group-based incentives in classroom contests can shift the competitive dynamics among students, potentially influencing their investments in human capital. Low-ability students, who typically rank lower in classroom assessments, may feel little incentive to invest in their education. As a result, they may become indifferent to their academic progress or even contribute negatively by disrupting the class. However, when placed in teams with higher-ability students, low-ability students gain a more equitable opportunity to be part of a winning team. This increased likelihood of success may motivate low-ability students to invest more effort in contributing to their team's success.

This notion is supported by past theoretical and empirical work on affirmative action and its effects on human capital investment. For example, Cotton, Hickman and Price (2020) presents a theoretical model predicting that disadvantaged students, particularly those of lower and middle ability, would increase their educational investments under affirmative action policies, as these policies reduce the competition they face. Conversely, high-ability students may reduce their investment under such policies. The authors further provide empirical evidence of these predictions in an experimental study with middle school children (Cotton, Hickman and Price, 2022).

Similarly, in the context of team contests, low-ability students may be motivated to increase their human capital investment as the social recognition of being part of a winning team becomes more attainable. High-ability students, motivated by the desire to win tournaments, also have an incentive to assist their lower-ability teammates, as the success of the team depends on the aggregate effort of all its members.

We implemented these ideas by designing EPInc – a digital platform for peer-driven knowledge diffusion – to proactively harness peer effects in classrooms. The platform achieves this goal with its three key features.

1. **Identify students' knowledge gaps**: Identifies students' knowledge gaps based on their homework performance and offers personalized learning content.

2. **Peer group manipulation**: Creates balanced teams by grouping students of different abilities. So, every team has high-ability students to help low-ability peers. Thus, high-ability

peers can help students with knowledge gaps and motivate them to study. High-ability peers may also set better group norms for work ethics for their teammates to emulate.

3. **Incentivize knowledge sharing**: Carefully cultivate team identities and hold team-based tournaments in classrooms. The social recognition of winning tournaments may motivate students of all abilities to help teammates and contribute to their team performance.

While helping teammates and investing individual efforts to contribute to their team's success, students may also develop character skills. Specifically, students could become more generous, reciprocal, and trusting toward their teammates and cooperate with them to secure high tournament ranks. Simultaneously, they may develop a competitive mindset toward students in other teams, reducing their social preferences for those outside their group. Therefore, group contests within classrooms have the potential to shape students' non-cognitive skills, such as generosity, trust, and reciprocity.

To evaluate the impact of fostering peer support through group contests on skill production, it is essential to compare this approach with scenarios where students are encouraged to self-learn and compete individually. Peer-driven knowledge sharing can address students' learning needs, as identified by personalized learning technologies, and could lead to higher and more equitable skill development within classrooms. Ideally, lower-ability students would benefit from interactions with their higher-ability teammates; however, unintended outcomes may also arise. For example, high-ability students might find collaborating with lower-ability teammates time-consuming and frustrating, leading them to feel that they have less time to focus on their studies. In these situations, encouraging students to self-learn in areas where they require improvement could prove more effective and may promote better overall skill development across the classroom. However, individual-based incentives could also exacerbate skill disparities within classrooms. While these incentives could motivate high-ability students to engage in self-learning, they might be less effective for low-ability students who need external support due to their below-grade-level understanding.

To evaluate the impact of peer-driven knowledge sharing through group contests on the EPInc platform, we compare it with two alternative approaches: (1) individual-based incentives combined with personalized learning and (2) traditional classroom methods of skill acquisition. Below, we outline the field study design aimed at achieving this objective.

3 Field Study Description

We introduced the EPInc platform in 40 grade 3 to 6 math classes in 9 schools in four Indian cities in August 2023. We added 46 additional grade 3 to 6 math classes in 4 schools in 2 Indian cities on the EPInc platform in January 2024. Of the 86 classes, 16 in 2 schools were all girls, 22 in 3 were all boys, and the remaining 48 in 10 schools were mixed gender. Each school in our experiment had at least two classes at each grade level.[5]

We randomly assigned classes in each school to the three categories – team, individual, and control groups – roughly in a 40:30:30 ratio. Students in the team and individual classes work on the EPInc platform, but control group students do not. As far as possible, care was taken to assign different groups across all grade levels.[6] Table 1 reports details of the number of schools, classes, and students in each group assignment.

[Table 1 about here.]

At the beginning of the experiment, we measured all students' baseline cognitive and non-cognitive skills. We measured students' cognitive skills with standardized tests prepared by educators outside the experimental schools based on the previous grade-level math syllabus. The student's performance on these tests measures their cognitive skills at the beginning of the experiment. We measured students' non-cognitive skills of generosity, reciprocity, and trust with social preference experiments. These non-cognitive skills were assessed based on students' responses in economic games, as detailed in Appendix C.

We formed teams in the team classes by balancing students based on their baseline cognitive abilities. Specifically, we randomly selected one student from each quartile of the baseline score distribution in a class to create the teams. After assigning four students to each team, the remaining students in the class were randomly distributed among the teams. As a result, each team consists of either four or five students. This random allocation of students to teams, based on the baseline score distribution within the class, ensures that the team formation is both external and balanced.

[5] We obtained approval (IRB#202201321) for our experimental procedure from a university's IRB board after multiple rounds of stringent checks

[6] An alternative to the within-school assignment was an across-school random assignment where all classes in a school were assigned to the same category. While this assignment mechanism minimizes the contamination of treatment across classes in a school (SUTVA violation), it introduces the effect of unobserved differences across schools in our estimates. We discuss how we minimize the effects of SUTVA violation in our setup in the later section

It's important to note that baseline scores may vary across different classes in various schools.

During the experiment, teachers provided class instructions on various math topics following the existing grade-level curriculum and pedagogical methods. After a few lectures, they assigned students homework (referred to as base homework) on the EPInc platform. Students in control classes completed their homework in the paper format within the classroom. In contrast, both individual and team students completed the base homework on the EPInc platform at school.

In the first part of the homework period, students answered homework questions. The platform identifies students' knowledge gaps (topics in which they did not perform well) based on their performance and provides feedback. While students in individual classes receive their topic-wise feedback, those in team classes receive topic-wise feedback for all team members. Appendix A.1 provides the screenshot of team feedback on the EPInc platform.

In the second part of the homework period, teachers show topic-wise feedback to a few randomly selected individual students on their computer tablets and encourage them to use this information to focus on their weaknesses in order to succeed in class tournaments. In team classes, teachers similarly selected a few teams to review the topic-wise performance of their members on tablets, encouraging students to assist their teammates in addressing their knowledge gaps to improve their chances in the tournaments. Apart from this feedback, students did not receive additional instructions or personalized help from the teachers. This study differs from others (Muralidharan, Singh and Ganimian, 2019; Banerjee et al., 2007) in which teachers provided personalized instructions and support during classes.

Teachers generate personalized homework on the EPInc platform for the team and individual classes a few days after the base homework. In this homework, students are offered more questions on topics in which students performed poorly in the base homework. Similar to the base homework, students do personalized homework in the assigned period in school. At the beginning of this period, teachers remind students to review their weak topics individually (with teammates) in individual (team) classes before they start the personalized homework. Students answer personalized homework questions in the second part of the period. For students in team classes, the last question of personalized homework asks them to select the teammate (or teammates) who has helped them with personalized homework. The students can also select the option that no one helped them. Control students are offered homework in place of personalized homework in their classroom. Thus,

the number of homework (school periods) given to the three categories of students was kept the same.

After a series of (base + personalized) homework assignments, teachers organize class tournaments. Teachers announce the tournament's schedule and the topics to be covered in advance, encouraging students to prepare individually (in individual classes) or with teammates (in team classes). During the tournament, students answer questions within a fixed time frame. In individual classes, students' tournament ranks are determined by their individual performance, while in team classes, team ranks are based on the aggregate performance of team members.

Following the tournament, teachers offer social recognition to the top three ranked teams (in team classes) or individuals (in individual classes). This recognition includes actions such as having the class clap for the winners and displaying their names on a social leaderboard in the classroom. The leaderboard, referred to as the "Champion Board," is shown in Appendix A.2. Since tournament scores do not contribute to students' final grades, social recognition serves as an incentive to motivate students to aspire to these achievements. In place of the tournament, control students receive either additional class instructions or extra homework assignments.

The cycle — consisting of class instructions followed by a series of (base + personalized) homework assignments and a class tournament — is repeated several times throughout the academic year in both team and individual classes. While the number of homework assignments and tournaments may vary across schools, teachers were instructed to provide an equal number of homework assignments and tournaments for all classes using the EPInc platform at the same grade level within a school. Efforts were made to ensure that all classes at the same grade level received a similar number of academic hours and math classes, as much as possible.

4 Data Description

4.1 Baseline Assessments

The EPInc platform contains a database of over 15,000 standardized multiple-choice questions covering all topics in the math syllabus of each grade level (grades 3 to 6) as prescribed by the National Council for Education Research and Training (NCERT) in India. The baseline test for each grade level was created by randomly drawing 30 questions on math topics covered in the syllabus

of the previous grade level from the database. Students' non-cognitive skills, including generosity, reciprocity, trust, and cooperative behavior, were measured using social preference experiments. Specifically, each student participated in economic games, such as the dictator and reciprocity games, with a randomly selected classmate.[7] These experiments provided baseline estimates of students' generosity, reciprocity, and trust. The baseline test was conducted for all students at the beginning of the academic session.

The schools in our sample cater to underprivileged children. About one-third of the students were absent from school on a given day. Therefore, we had to perform multiple rounds of baseline tests. Finally, we could cover 90 percent of the total students (2810 of 3110 students).[8] After assigning students with baseline scores in our sample, we randomly assigned those without scores to different teams in team classes. As a result, while most teams comprised four students, some had five students.

Next, we compare the baseline scores of the students in the three groups in Table 2. We provide standardized baseline scores (normalized to a mean of zero and a standard deviation of one) for cognitive and non-cognitive tests for the three groups of students. Students' non-cognitive scores for generosity, reciprocity, and trust are highly correlated. Therefore, we use the composite non-cognitive scores as predicted by the principal component analysis as a measure of students' social preference for others in our analysis. Table 2 reports the summary statistics.

[Table 2 about here.]

In Section 5.2, we check if the three categories of students are balanced in their cognitive and non-cognitive skills at the beginning of our intervention.

4.2 Homework and Tournaments

Students build on their initial math skills by doing base and personalized homework and preparing for tournaments during the experiment. We provided detailed schedules for homework and tournaments at schools. Typically, teachers were required to conduct a tournament after two sets of base and personalized homework. We ensured that all teams and individual classes at the same grade

[7] Details of social preference experiments are available in Appendix C.

[8] We show that the attrition in students across the three types of classes was statistically similar.

level within a school received an equal number of homework assignments and tournaments on the EPInc platform. However, the schedule was sometimes disrupted due to unforeseen school closures or other activities. Additionally, students missed homework assignments and tournaments if they were absent from school on the designated days.

We collect data on student attendance and scores for both homework and tournaments in team and individual classes. Table 3 shows the total number of homework assignments offered and attended, along with the average scores achieved by the two student groups. On average, students in individual classes attend 11.3 out of 16.6 (68%) of the homework assignments offered, while those in team classes attend 11.1 out of 16.3 (68%) of the offered homework. Overall, students in our sample were provided with over 1.2 million math questions across approximately 60,000 homework assignments during the experiment. The attendance and performance in homework for the two groups of students were statistically similar, with the average standardized homework score for both team and individual classes being approximately 0.21.[9]

[Table 3 about here.]

While the homework schedule was not explicitly communicated to students, the tournaments were announced in class. On average, students in individual classes attended 2.5 out of the 3.8 tournaments offered (64%), while students in team classes attended 2.5 out of 3.9 tournaments (65%). Although both types of students participated in a similar number of tournaments, their average scores differed significantly. Students in team classes achieved an average standardized score of 0.16, while those in individual classes scored an average of -0.18. This difference is statistically significant at α=0.01. Since students answer questions drawn from a standardized question bank (created outside their school) within a fixed time during tournaments, these scores are also a measure of their skill levels.

4.3 Endline Assessments

We conducted a paper-based endline assessment for all students at the conclusion of the academic session. Endline tests were administered in February 2024 for schools that began in August 2023 and in December 2024 for those that started in February 2024. Question papers for the endline

[9] results available on demand from the authors.

assessment were prepared by experts outside of the schools involved in our experiment and consisted of standardized multiple-choice questions based on the math syllabus for each grade level as prescribed by the National Council for Educational Research and Training (NCERT) in India. The choice of a paper format for the assessment aimed to provide an equal opportunity for students in control classes who did not use tablets during the academic session.

We conducted social preference experiments to assess students' non-cognitive skills at the end of the study. Students participated in economic games with randomly selected classmates from individual and control classes. In team classes, however, half of the students were randomly assigned to play economic games with classmates outside their team, while the other half played with their randomly chosen teammates. This setup allows us to evaluate the non-cognitive skills of students in teams toward their teammates as well as toward classmates in other teams. Summary statistics can be found in Table 4.

[Table 4 about here.]

5 Empirical Model and Results

5.1 Skill Production in Schools

Control students learned math through standard school practices, including class instruction, homework, and studying for assessments. Our intervention provided topic-specific feedback to students based on their homework performance. Individual students received feedback on their own performance, while team students received feedback that included the performance of all their team members. We implemented incentives, such as social recognition for winning tournaments, to encourage students to address their knowledge gaps based on this feedback. Individual students could use the feedback to self-learn the topics they struggled with. In contrast, team students could use the feedback not only to self-learn but also to assist and seek help from their teammates. Teachers did not offer personalized instructions or support to students in either the team or individual classes. Control students did not receive personalized learning or participate in any tournaments.

Finally, students' math skills were assessed by the standardized endline assessments conducted at the end of the experiment. We examine the relative efficacy of skill production in team and

individual classes compared to control classes with the following econometric specifications:

$$EndlineScore_i = \beta_1 TeamGr_i + \beta_2 IndividualGr_i + \beta_3 BaselineScore_i + \alpha_s + \alpha_g + \epsilon_{it}; \quad (1)$$

where i denotes the unit of analysis. We conduct our analysis at (1) the class level at which we randomly assign the team and individual treatments on the EPInc platform and (2) the student level. Variables $EndlineScore_i$ and $BaselineScore_i$ denote the endline and baseline scores (cognitive and non-cognitive) of unit i (normalized to μ=0 and σ=1). Variables $Team_i$ and $Individual_i$ are the indicator variables for student i belonging to the team and individual classes, respectively.

We include school and grade fixed effects (α_s and α_g) to account for unobserved differences in educational resources, environment, and students across schools and grades. We cluster-correct standard errors at the class level, at which the treatment varies. For robustness, we also report the p-values from standard errors cluster corrected at the school level. Given the small number of classes (86) and schools (13), we also use the wild cluster bootstrap-t method of Cameron, Gelbach and Miller (2008). Finally, we also computed the p-values from a permutation test by randomly shuffling the treatment at the class level.

The coefficient of $IndividualGr$ in Equation (1) identifies the treatment effect of individual feedback and self-learning over traditional skill production in control classes. The $TeamGr$ coefficient identifies the treatment effect of team feedback and team help compared to skill production in control classes. Additionally, we also compare the skill production in team and individual classes.

To identify unbiased treatment effects, we must ensure that the main assumptions of average treatment effects are adequately met in our experiment. First, let's discuss the SUTVA assumptions from both the recipients' (students) and providers' (teachers) perspectives. Access to the EPInc for treated classes should not influence the potential outcomes of control students. For example, the control students could interact with the team students and learn about the EPInc platform. The institutional arrangements limit the possibility of interaction between students in treated and control classes, as students typically attend all classes with the same classmates. Their interactions with students from other classes occur only during lunch, and that too only if their lunch periods

coincide.[10] Students benefit from using EPINc by receiving feedback and sharing knowledge with their peers. In contrast, control group students can only learn about EPINc without actively using it, so their skill development is unlikely to be impacted. There is a scenario where simply knowing about EPINc could influence learning; for instance, students might feel demotivated to engage in their education due to a lack of access to the EPINc platform. However, this scenario is unlikely, as students can only answer math questions on the EPINc platform while under supervision at school.

Many teachers instruct multiple classes, which means they may be teaching both EPInc-based treatment classes and control classes within the same school. The reduced effort required for creating and grading homework on the EPInc platform might affect their management of paper-based homework in the control classes. As a result, the potential learning outcomes for students in the control group could be influenced by the EPInc treatment.

To address this issue, we encouraged teachers to include questions from the EPInc question bank in the paper-based homework to control students. This approach not only minimizes their workload for administering homework in control classes but also ensures that all students engage with similar-quality homework questions. Additionally, we monitored to ensure that the number of academic hours and activities offered to the control classes at each grade level was comparable to those provided in the treatment classes.

Our treatment effect estimates infer the potential outcomes for treated students under no treatment from the outcomes of control students. We test the validity of this assumption by conducting balance checks in the following section.

5.2 Balance check

To identify the effect of two treatments in our setup, it's essential that the students assigned to different groups begin with comparable skills, both cognitive and non-cognitive. We check this fact with the following specifications: (2).

$$BaselineScore_i = \beta_1 Team_i + \beta_2 Individual_i + \alpha_s + \alpha_g + \epsilon_i; \tag{2}$$

[10] It's important to note that, unlike in U.S. schools—where students move to different classrooms depending on the subject—they remain with the same classmates throughout the day in Indian schools, while teachers come to the students' classrooms to deliver lessons.

where $Team_i$ and $Individual_i$ are the indicator variables for student i belonging to the team and individual classes, respectively. The dependent variable $BaselineScore_i$ denotes student i's cognitive/non-cognitive score in the baseline test (normalized to μ=0 and σ=1). Note that the initial skill levels for students may differ across schools and grade levels. We account for these differences with school and grade fixed effects (α_s and α_g) in our specification. Since all students in a class are tested together, unobserved environment- and time-related factors in a class may affect their baseline scores. We account for this possibility by cluster-correcting the standard errors at the class level.

[Table 5 about here.]

Columns (1) to (2) of Table 5 show estimates for students with (1) baseline scores and (2) both baseline and endline scores from Equation (2). Insignificant estimates for $Team$ and $Individual$ indicators suggest a lack of statistical difference in baseline math skills across students from the three groups.

We aim to understand how peer effects differ among students of varying abilities in classrooms. To do this, we categorized students into three groups based on their baseline scores: low ability (below the 33rd percentile), medium ability (between the 33rd and 66th percentiles), and high ability (above the 66th percentile). We then estimated Equation 2 separately for each ability group. The estimates for the coefficients in Columns (3) to (5) of Table 5 show that the variables for $Team$ and $Individual$ are statistically insignificant across all specifications. This indicates that the baseline cognitive scores among the three groups are statistically indistinguishable in each ability segment.

We also examine the balance of non-cognitive abilities of students in the three categories of classes. To do this, we compare the composite non-cognitive scores of the three categories of students in Equation (2) and report resulting estimates in Table 6. Our findings indicate that non-cognitive skills are well-balanced among all students, as well as within each ability tercile, across the three categories of classes.

[Table 6 about here.]

5.3 Cognitive Skill Production Estimates

5.3.1 Class-Level Cognitive Skill Production Estimates

Students in team and individual classes obtain personalized learning with incentives to perform well in tournaments, they should acquire higher skills than students in control classes. Furthermore, if students share knowledge with their teammates from team-based feedback, it could produce higher and more uniform skills in team classes than in individual classes. We examine these possibilities by comparing the average values and distributions of endline scores in the three types of classes.

We computed the average value and Coefficient-of-Variation (COV) for baseline and endline cognitive scores for each class in our sample. We further computed the average scores for only high, medium, and low-ability students in each class. Since the classes in our experiment were randomly assigned to team, individual, and control classes, we found insignificant estimates for $TeamGr$ and $IndividualGr$ coefficients from Equation 2 on class-level variables [11]. This indicates statistically indistinguishable baseline skill distributions across the three categories of classes. Next, we compare the endline skill distributions among these classes in Table 7 by estimating Equation (1) on class-level variables.

[Table 7 about here.]

Column (1) of Table 7 shows a positive and significant estimate for $TeamGr$ but not for $IndividualGr$, indicating statistically higher (similar) average endline scores in team (individual) classes than control classes. The estimated coefficient value suggests that team students obtained 0.23σ higher endline scores than control students. The negative and significant estimates of $TeamGr$ in column (2) indicate that team classes have a lower Coefficient-of-Variation for endline scores than in control classes, suggesting more uniform learning in team classes. Finally, the coefficients $TeamGr$ in columns (3) and (4) suggest that high and medium-ability team students obtain 0.22σ and 0.33σ higher end-line scores than the corresponding control students. The coefficient of $TeamGr$ for low-ability students is positive and insignificant (at $\alpha = 0.11$).

Thus, we find evidence for higher and less dispersed skill production in team classes than in control classes, but similar learning outcomes for individual and control classes.

[11] Estimates available on demand from the authors

5.4 Student Level Cognitive Skill Production Estimates

Next, we examine the intent-to-treat (ITT) effect of team and individual assignment on skill production using student-level cognitive scores in Equation (1). The coefficients β_1 and β_2 identify the ITT of team and individual-based skill production over the control classes.

[Table 8 about here.]

Columns (1) to (4) of Table 8 report the coefficient estimates from Equation (1) for all students, high, medium, and low-ability students separately. We find positive and significant estimates for variable $TeamGr$ in all categories but not for $IndividualGr.$, indicating higher tournament scores for students in team classes than control classes but not for individual classes. The magnitude of the estimated coefficient suggests that students in team classes obtain 0.25σ higher endline scores than their counterparts in control classes.

Next, we estimate Equation (1) on students from team and individual classes to examine how team-based skill production compares with individual-based skill production on the EPInc platform in Columns (5) to (8) of Table 8. We find positive and significant estimates for team-based skill production for high and medium abilities but not low-ability students.

5.5 Mechanism

The EPInc platform is designed to improve skill production with the following mechanisms.

1. *Feedback* - The platform provides team members feedback about each other's knowledge gaps. Students can utilize this information to render or seek help from their teammates.

2. *Induce peer help* - Knowing each others' knowledge gaps and motivation to win tournaments may induce students to share knowledge with their team members.

3. *Improve learning motivation* - Social recognition from winning tournaments can motivate students to invest in learning from the knowledge shared by peers.

In the following sections, we provide suggestive empirical evidence for these mechanisms.

5.6 Feedback Effect

In Appendix A.1, we describe how the EPInc platform provides students feedback on their knowledge gaps based on their homework performance. While students in individual classes obtain feedback on their homework performance, those in team classes receive information about the knowledge gaps of all team members. Indeed, self-feedback can inform students which areas to invest effort in to improve their overall learning. Team feedback informs students not only about their knowledge gaps but also about the strengths and weaknesses of their team members. Team students can, of course, invest efforts to self-learn in their knowledge gaps. But if they cannot, they know whom to ask for help in their team. At the same time, teammates receive information about students' knowledge gaps and are urged to help them. Thus, team feedback provides higher opportunities for learning than self-feedback in individual classes. Control students do not use the EPInc platform, and thus, they do not obtain any feedback. [12]

Individual students obtain feedback on their knowledge gaps (self-feedback) after completing the homework on the EPInc platform. Thus, we measure self-feedback with the proportion of homework completed by students in individual classes ($SelfFeedback$). In contrast, team feedback for team students depends on how many team members attended homework with them. Therefore, we computed the total number of team members (including the student) attending each homework and then averaged this value over all offered homework. Thus, we measure the team feedback a team student receives ($TeamFeedback$) with the average number of teammates (including the student) attending homework with her. In our data, $SelfFeedback$ has a mean value of 0.68 (SD=0.26), indicating individual students, on average, attend 68% of homework. Variable $TeamFeedback$ has a mean value of 2.15 (SD=1.01), showing, on average, 2.15 team members attend homework in team classes. The large standard deviation values for these variables suggest variations in feedback across students that help us identify its effect on skill production.

We examine the relationship between self (team) feedback received by students in individual (team) classes with their baseline scores in Appendix B. We find a positive significant relationship between self-feedback with students' cognitive abilities in individual classes, suggesting it is endoge-

[12] Control students may get delayed feedback from their graded paper homework, as teachers take time to grade homework. Thus, our estimate would identify the relative efficacy of Edtech-generated instantaneous personalized feedback over traditional delayed paper-based feedback.

nous to students' potential learning outcomes. However, we did not find a significant relationship between students' team feedback and baseline scores in team classes, suggesting no relationship between team attendance and student's baseline abilities. However, homework attendance is not randomly assigned, and thus, we instrument self-feedback and team feedback variables with the random assignment variable for individual and team classes ($IndividualGr$ and $TeamGr$). Accordingly, we estimate the following specifications to test the effects of self and team feedback on students' cognitive skill production:

$$\begin{aligned} EndlineScore_i = \beta_1 TeamFeedback_i + \beta_2 SelfFeedback_i + \\ \beta_3 BaselineScore_i + \alpha_s + \alpha_g + \epsilon_i; \end{aligned} \tag{3}$$

The variables $TeamFeedback$ and $SelfFeedback$ are equal to zero for students in control classes. First, we report the IV estimates from Equation (3) for all students in Table 9. We further check if feedback effects vary across students of different abilities, with separate IV models for the high, medium, and low-ability students. The IV estimates identify the average causal response (ACR) – a weighted average response of an additional self/team feedback on skill production for those whose treatment status is affected by the instruments (Angrist and Imbens, 1995).

[Table 9 about here.]

We find positive and significant coefficients for $TeamFeedback$ in columns (1) and (2), indicating that the team students with higher team feedback obtain higher endline scores than the control students. Specifically, the estimate estimated coefficient value means that additional teammate attendance can improve the students' endline scores by 0.11σ. We also find that team feedback helps students of all abilities. The high, medium, and low-ability students benefit by 0.12σ, 0.12σ, and 0.09σ. However, we fail to find evidence that self-feedback results in higher endline scores for individual students than control students, suggesting that self-feedback does not result in higher skill production for students.

We cannot reject the equality of the estimates from our IV and OLS models (Columns 1 and 5 in Table 9), which suggests the similarity between LATE and ATE estimates. We further estimate the OLS model of Equation (3) for all and only (team + individual) students (Columns (5) and (6) in Table (9)). The constant terms, indicating the average outcome for students not receiving

the treatment in these samples, in the two columns are statistically indistinguishable. Since the students not receiving treatment are (control students + non-compliers) in the full sample and only non-compliers in the (team + individual) student samples, it indicates the equality of potential outcomes across different compliance groups. Perhaps for this reason, we find similar estimates in Equation's (3) OLS and IV models.

Our results of higher skill production in team students from the higher team feedback could be due to knowledge sharing among team members. Interestingly, no effect of self-feedback on skill production of individual students could be due to their lack of motivation or inability to self-learn without support.

5.7 Team Interactions Effect

The feedback about knowledge gaps only helps students if they act upon it. Thus, interaction among team members based on feedback should help students. We further use information about help received (given) from (to) teammates on the EPInc platform to examine the effect of team interactions on students' skill production.

The last question on personalized homework on the EPInc platform asks team students to indicate the teammate(s) who helped them with personalized homework. Students also have the option to say that no one helped them. [13] Thus, we capture data on who sought help from whom in each homework in team classes. We discuss different EPInc screens that nudge students to share knowledge with their teammates in Appendix A.3.

Of the 1,196 team students, 1046 sought help from or gave help to teammates in at least one homework. We computed the total team help a team student received in homework as the sum of the number of teammates who received help from her (as reported by them), helped her (as reported by the student), and self-help. We include self-help in the team help because the student can also self-study based on her homework feedback besides her teammates' help. Then, we compute the average team help for a team student ($AvgTeamHelp$) as the sum of team help in homework she attended to the total number of offered homework. Students in individual classes can only help themselves based on their homework feedback. Consequently, we calculate the average self-help an individual student received in homework ($AvgSelfHelp$) as the ratio of attended to offered

[13] Individual students are not asked this question in personalized homework

homework.[14] The $AvgTeamHelp$ is the actual help received/given instead of total team attendance in the $TeamFeedback$ variable in Section 5.6. In our data, $AvgTeamHelp$ has a mean=0.98 (SD=0.46) – 0.33 (SD=0.26) due to interaction with teammates and 0.66 (SD=0.28) due to attending offered homework. This means that team students, on average, interact with 0.33 teammates per homework and attend 66% of the offered homework. The $AvrSelfHelp$ variable has a mean value of 0.68 (SD=0.26), suggesting individual students, on average, attend 68% of homework.

We use the above-computed variables to estimate the effect of reported team interactions on students' skill production. Naturally, team interactions in team classes and self-study in individual classes may be endogenous to their potential learning outcomes. Therefore, we estimate the LATE estimate for the effect of these variables on students' endline scores by instrumenting them with the random team and individual assignment variables in the following specifications:

$$\begin{aligned} EndlineScore_i = \beta_1 AvgTeamHelp_i + \beta_2 AvgSelfHelp_i + \\ \beta_3 BaselineScore_i + \alpha_s + \alpha_g + \epsilon_i; \end{aligned} \tag{4}$$

The variables $AvgTeamHelp$ and $AvgSelfHelp$ are equal to zero for students in control classes. All other variables have the same meaning as in previous specifications. We estimate the IV model of Equation (4) for all students and students in each ability tercile separately. We further estimate the OLS model of Equation (4) for all and (team + individual) students to explore if the potential outcome for non-compliers is similar across different student groups. Table 10 reports these estimated coefficients from Equation (4).

[Table 10 about here.]

We find a significant estimate for $AvgteamHelp$ for all students in Column (1) of Table 10, indicating that an additional team help per homework improves students' endline scores by 0.24σ. However, the insignificant effect of $AvgSelfHelp$ indicates the failure to find an effect of self-help on students' endline scores.[15] Team students of all abilities obtain higher endline scores than the control students. Columns (5) and (6) provide OLS estimates on all and (team + individual) students. The constant terms, indicating the average outcome for students not receiving the treatment, in

[14] Note that $AvgSelfHelp$ is the same as $SelfFeedback$ variable in Section 5.6.

[15] The $AvgSelfHelp$ variable is a noisy measure of self-help – it captures the self-feedback students received but not whether they self-studied

the two columns are statistically indistinguishable. Since the students not receiving treatment are (control students + non-compliers) in the full sample and non-compliers only in the (team + individual) student samples, it indicates the equality of potential outcomes across different compliance groups.

We find positive and significant estimates for $AvgteamHelp$ in columns (2) to (4) of Table 10, suggesting team help benefits students of all abilities. These results indicate that while interacting with high-ability teammates is beneficial for low-ability students, interacting with low-ability teammates is also beneficial for high-ability students. The literature on learning by teaching suggests that high-ability students benefit from tutoring their teammates (Guerrero and Wiley, 2021; Fiorella and Mayer, 2013; Shah, N.d.).

5.7.1 Topic-wise Score Analysis

The team members can help students with their weak topics if they know them well. If this is true, team students should do well in topics where their teammates have a higher capacity to help them – i.e., they have obtained high homework scores on those topics. To test this fact, we first classified all math topics in our data into nine major topics (such as geometry, data interpretation, and measurement) and computed topic-wise students' homework, tournament, and endline scores.

It is plausible that topic-wise homework scores of students in a class may be correlated because they learn from the same teacher. However, team students additionally learn topics from interacting with their teammates who understand them well. Thus, team students' learning on a math topic will be correlated to their classmates' skills as well as their teammates' skills on that topic. For each student in the team and individual classes, we computed classmates' topic-wise math skills by dividing the sum of topic-wise homework scores of classmates (leaving out the student) by the number of her classmates ($OtherClassHWScore$). For the team students, We computed the teammate's topic-wise math skills by dividing the sum of topic-wise homework scores of teammates (leaving out the student) by the number of her teammates ($OtherTeamHWScore$). The variable $OtherTeamHWScore = 0$ for individual students because they work individually.

Next, we examine if team students perform better in math topics in which their teammates

have performed well in the homework with the following specifications:

$$Y_{ij} = \beta_1 OtherTeamHWScore_{ij} + \beta_2 OtherClassHWScore_{ij} + \beta_3 BaselineScore_i + \alpha_s + \alpha_g + \alpha_j + \epsilon_{ij}; \tag{5}$$

where i and j denote students and math topics. We estimate Equation (5) for two dependent variables: (1) the average tournament score and (2) the endline score of student i in topic j. The independent variables $OtherTeamHWScore$ is the average homework score of the student's teammate and $OtherClassHWScore$ is the average homework score of i's classmates in topic j. We also add the topic fixed effects (α_j). We estimate both the OLS and IV models and report the resulting estimates in Table 11.

[Table 11 about here.]

In all columns of Table 11, we find positive and significant estimates for coefficient β_2 that capture the correlation between the student's and her classmates' scores on a topic. These results indicate intra-class correlations in topic-wise assessment scores for students. However, we also find positive and significant estimates for coefficient β_1, indicating an additional effect of team members' higher homework scores on a topic on student's assessment scores on that topic – i.e., team students obtained higher endline/tournament scores in topics in which their teammates obtained high homework scores. This provides another evidence that students learn from team members' help in team classes.

5.8 Team Rank Effect

Winning tournaments can provide social recognition, which may motivate students to study. In individual classes, only the top three students received recognition, whereas in team-based classes, the top three teams—comprising 12-14 students, including those with medium and lower abilities—were acknowledged. As a result, more team students may be incentivized to seek help from teammates and study to contribute to their team's success.

Individual tournament rankings reflect students' personal performance and may correlate with their baseline cognitive scores. In contrast, team rankings are based on aggregate scores, and since

teams are balanced in baseline abilities, they have equal chances of winning. This suggests that a team student's tournament rank is not necessarily linked to the average cognitive abilities of their teammates.

To analyze this, we calculated the average percentile rank each student achieved in tournaments across both team and individual classes. Students in team-based classes were assigned HighTeamRank = 1 if their average percentile rank was above the class median; otherwise, they were assigned 0. We then examined the relationship between students' likelihood of obtaining high tournament ranks and their baseline cognitive scores using the following specifications:

$$HighTeamRank_i = \beta_1 BaselineScore_i + \alpha_s + \alpha_s + \epsilon_{it}; \tag{6}$$

We estimate Equation (6) for the samples of (1) individual students, (2) team students, (3) team students with high tournament rank ($HighTeamRank$ =1) with individual and control students, and (4) team students with high tournament rank with individual students. Note that variable $HighTeamRank$=0 for individual and control students. Table 12 reports the estimates.

[Table 12 about here.]

Column (1) of Table 12 reports a positive significant estimate for $BaselineScore$ for individual students, suggesting that those with higher cognitive scores are more likely to achieve high tournament ranks in individual classes. This aligns with expectations, as rankings in these classes are based on individual performance. In contrast, columns (2) to (4) report insignificant estimates for $BaselineScore$, no statistically significant differences in baseline scores across the following comparisons: (1) team students with high versus low tournament ranks, (2) team students with high tournament ranks versus individual and control students, and (3) team students with high tournament ranks versus individual students. The intuition behind these results is that teams inside classes are balanced in their members' cognitive abilities, and the team's tournament ranks are not correlated to individual students' efforts/abilities but the whole team's effort vis-a-vis other teams' efforts in the class.

Team students who achieve high tournament ranks, despite having similar baseline cognitive abilities as others, may be more motivated to study and share knowledge with teammates. This

suggests they could develop higher skills during the experiment. To test this, we compare the skill acquisition of team students securing high and low tournament ranks with individual and control students. We define two indicator variables: $HighTeamRank = 1$ for team students whose average percentile tournament rank is above the class median (0 otherwise), and $LowTeamRank = 1$ for those below the median (0 otherwise). Students in individual and control classes are assigned zero for both variables. We then estimate the following specifications to assess the impact of students' tournament ranks on their endline scores.

$$\begin{aligned} EndlineScore_i = \beta_1 HighTeamRank_i + \beta_2 LowTeamRank_i + \\ \beta_3 BaselineScore_i + \alpha_s + \alpha_g + \epsilon_{it} \end{aligned} \tag{7}$$

For unbiased estimates of β_1 and β_2 in Equation (7), $HighTeamRank$ and $LowTeamRank$ should be exogenous to students' potential outcomes. Statistically similar baseline cognitive scores for team students obtaining high and low team ranks with each other and individual and control students in Table 12 supports this possibility. Accordingly, we estimate the OLS model for Equation (7). The coefficients β_1 and β_2 identify the endline scores of team students obtaining high and low team ranks compared to the full sample of individual and control students. Considering that team ranks are not randomly assigned, we also estimate the IV model by estimating Equation (7) with only the $HighTeamRank$ variable and instrumenting it with the team assignment variable.

[Table 13 about here.]

Table 13 reports the resulting estimates. We find a positive significant estimate for $HighTeamRank$ but an insignificant estimate for $LowTeamRank$ in column (1), indicating that team students with high tournament ranks obtain higher endline scores than individual and control students, but those securing lower ranks obtain statistically similar endline scores. The results in column (2) remain similar when we estimate this model for the sample of (team + individual) students. Columns (3) and (4) show the IV estimates for all and only (team + individual) classes. The estimated coefficients from IV specifications are larger but have the same sign and significance.

Higher tournament ranks could motivate students to study more (higher effort) and have higher academic interaction with teammates (peer effects), which is reflected in their higher endline scores. It is also possible that better team dynamics – higher interactions and team members' efforts –

could result in higher skill production, which is reflected in higher tournament ranks and endline scores. However, crucial ingredients in both of these paths for higher skill production are higher students' efforts to learn and higher help offered to them by team members. [16]

We estimate Equation (7) separately for high, medium, and low-ability students to examine the relative effect of team ranks across them. Table 14 reports the estimates. We find a positive significant effect of high tournament ranks for high and medium-ability students but not for low-ability students. These results suggest that high tournament ranks lead to higher skill production in high and medium-ability team students, perhaps because it motivates them to invest higher support to their team members.

[Table 14 about here.]

5.9 Non-cognitive Skill Production Estimates

Next, we examine how students' social preferences for their peers change due to competing in tournaments individually and in groups compared to those in control classes in traditional classrooms. Specifically, we estimate the following econometric specifications:

$$\begin{aligned} EndlineScore_i = \beta_1 TeamGrPeer_i + \beta_2 TeamGrNoPeer_i + \beta_3 IndividualGr_i \\ + \beta_4 BaselineScore_i + \alpha_s + \alpha_g + \epsilon_{it}; \end{aligned} \tag{8}$$

where $TeamGrPeer$ and $TeamGrNoPeer$ are the indicator variables for team students' non-cognitive skills for a random teammate and classmate from another team. Variables $EndlineScore$ and $BaselineScore$ denote the predicted non-cognitive scores by combining students' generosity, reciprocity, and trust scores in the principal component analysis. The remaining variables have the same meaning as in previous specifications. Table 15 reports the estimates of Equation (8) for all, high, medium, and low-ability students.

[Table 15 about here.]

The insignificant estimate for $IndividualGr.$ in Column (1) indicates statistically similar students' non-cognitive skill production (social preference formation) in individual and control classes.

[16] Note that students can only benefit from the knowledge shared by their teammates if they put in the effort to learn

However, the differences in non-cognitive skills of team and control students are more nuanced. While team students demonstrate comparable social preferences toward their teammates, they exhibit significantly lower preferences for students in other teams compared to control students (as indicated by the insignificant $TeamGrPeer$ estimate but the negative significant $TeamGrNoPeer$ estimate in Column (1)). This suggests that working in teams fosters stronger competitive attitudes toward students in other teams while maintaining stable preferences for teammates. Moreover, this competitive dynamic is primarily driven by high-ability team students rather than their lower-ability counterparts. Notably, the negative significant coefficient for $IndividualGr$ among high-ability students indicates that these students develop reduced preferences for their classmates compared to control students. Overall, these findings suggest that high-achieving students experience heightened competitive feelings toward others when engaging in digital tournaments.

6 Discussions and Conclusion

We present experimental evidence that proactively designing peer interactions inside classrooms can significantly improve skill production in K-12 classrooms. Our software platform encourages students to share knowledge with teammates in order to win class tournaments, and we implemented it in multiple classrooms across Indian schools to assess its impact. Results show that students collaborating in teams achieved 0.25σ and 0.17σ higher scores in endline tests compared to those who worked individually on the platform or had no access to it. However, there was no statistically significant difference in learning outcomes between students incentivized to self-learn on the platform and those without access. Notably, skill gains from team-based learning were consistent across all ability levels, dispelling concerns that interactions with lower-ability peers might hinder high-ability students' progress.

Our intervention showcases the potential of technology to identify students' knowledge gaps and induce peer-driven knowledge sharing with group-based incentives in classrooms. Students achieve higher scores when more teammates receive feedback on their learning gaps while doing homework with them. Access to information about peers' learning deficits enables more effective knowledge sharing. We observed that students who engaged more frequently with their teammates performed better on endline tests. Additionally, the social recognition gained from achieving high tournament

ranks appears to motivate students to seek/offer help from/to teammates and increase their learning efforts. Notably, students in high-ranking teams also achieved higher individual endline scores, suggesting that a strong sense of team identity drives greater investment in education and peer support. To maximize these benefits, EdTech applications should design information displays that reinforce team identity and provide social recognition for tournament success. Strengthening students' motivation to win team contests can significantly enhance academic engagement and peer-driven learning.

Our findings show that incentivizing peer interactions can leverage peers – an underutilized resource in the K-12 ecosystem – to improve skill production within existing school resources and setups. Our team-based skill production on the EPInc platform improves math scores by 0.25σ. Students engaged with the platform for approximately 200 minutes per month (65 minutes per week of school time), including base homework, personalized homework, and tournament participation. [17] Even this limited-time engagement on the platform effectively motivated students to share knowledge with teammates, leading to significant improvement in skill acquisition. The experiment, implemented with 3,100 students, incurred a monthly cost of INR 400 (USD 5) per student, covering software development, IT infrastructure, and staffing. Compared to other educational interventions in Indian schools, our approach delivers similar skill gains while requiring significantly less student time on the platform, minimal curriculum or pedagogical adjustments, and substantially lower costs (Muralidharan, Singh and Ganimian, 2019; Berry and Mukherjee, 2016). Even when prorating the high fixed costs across a relatively small student base, the EPInc platform yields greater returns per unit of time and money spent, highlighting its cost-effectiveness and scalability.

Although our findings appear to contradict the policy of tracking students to classrooms based on their abilities, it actually complements the tracking policy. While the primary benefit of tracking policy comes from the teacher's ability to give instructions at an appropriate level, such instructions could still result in heterogeneous learning across students. For instance, students of similar abilities may differ in their grasp of different concepts taught in the class. Organizing team contests in such a setting may promote knowledge sharing and, more importantly, motivate students to invest in learning activities to improve skill production.

[17] students spend 2 x (40 minutes base homework + 40 minutes personalized homework) + 40 minutes tournament per month = 200 minutes per month or 65 minutes of school time on the platform per week.

Our platform can easily integrate other resources in the K-12 ecosystem, further enhancing skill production in K-12 schools. Once students' knowledge gaps are identified, we can connect them with other resources for help. For instance, informing parents about their child's learning deficits and providing relevant educational resources empowers them to assist directly or seek external help. Additionally, the platform offers real-time, topic-wise aggregate class performance data, enabling teachers to adjust their instructional strategies accordingly. If learning gaps stem from ineffective classroom instruction, our platform can incorporate external educators and mentors to support both students and teachers. Beyond student-focused interventions, the platform can also facilitate collaborative professional development for educators. Just as team contests drive peer learning among students, similar competitions could be designed for teachers to foster knowledge-sharing and instructional improvement.

Overall, integrating different resources on an EdTech platform would allow them to complement each other to maximize skill production. Future research should examine creative ways to harness these complementarities on the EdTech platform.

References

Angrist, Joshua D and Guido W Imbens. 1995. "Two-stage least squares estimation of average causal effects in models with variable treatment intensity." *Journal of the American statistical Association* 90(430):431–442.

Banerjee, Abhijit V, Shawn Cole, Esther Duflo and Leigh Linden. 2007. "Remedying education: Evidence from two randomized experiments in India." *The quarterly journal of economics* 122(3):1235–1264.

Berry, James and Priya Mukherjee. 2016. "Pricing of private education in urban India: Demand, use and impact." *Unpublished manuscript. Ithaca, NY: Cornell University* .

Betts, Julian R and Andrew Zau. 2004. "Peer groups and academic achievement: Panel evidence from administrative data." *Unpublished manuscript* .

Betts, Julian R and Jamie L Shkolnik. 2000. "The effects of ability grouping on student achievement and resource allocation in secondary schools." *Economics of Education Review* 19(1):1–15.

Bishop, John. 2006. "Drinking from the fountain of knowledge: Student incentive to study and learn–externalities, information problems and peer pressure." *Handbook of the Economics of Education* 2:909–944.

Bowles, Samuel and Herbert Gintis. 2011. A cooperative species: Human reciprocity and its evolution. In *A cooperative species*. Princeton University Press.

Bulman, George and Robert W Fairlie. 2016. Technology and education: Computers, software, and the internet. In *Handbook of the Economics of Education*. Vol. 5 Elsevier pp. 239–280.

Burke, Mary A and Tim R Sass. 2013. "Classroom peer effects and student achievement." *Journal of Labor Economics* 31(1):51–82.

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

Carrell, Scott E, Bruce I Sacerdote and James E West. 2013. "From natural variation to optimal policy? The importance of endogenous peer group formation." *Econometrica* 81(3):855–882.

Carrell, Scott E, Richard L Fullerton and James E West. 2009. "Does your cohort matter? Measuring peer effects in college achievement." *Journal of Labor Economics* 27(3):439–464.

Choi, Jung-Kyoo and Samuel Bowles. 2007. "The coevolution of parochial altruism and war." *science* 318(5850):636–640.

Cotton, Christopher, Brent R Hickman and Joseph Price. 2020. Affirmative action, shifting competition, and human capital accumulation: A comparative static analysis of investment contests. Technical report Queen's Economics Department Working Paper.

Cotton, Christopher S, Brent R Hickman and Joseph P Price. 2022. "Affirmative action and human capital investment: Evidence from a randomized field experiment." *Journal of Labor Economics* 40(1):157–185.

Dechenaux, Emmanuel, Dan Kovenock and Roman M Sheremeta. 2015. "A survey of experimental research on contests, all-pay auctions and tournaments." *Experimental Economics* 18:609–669.

Duflo, Esther, Pascaline Dupas and Michael Kremer. 2011. "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya." *American economic review* 101(5):1739–1774.

Epple, Dennis and Richard E Romano. 2011. Peer effects in education: A survey of the theory and evidence. In *Handbook of social economics*. Vol. 1 Elsevier pp. 1053–1163.

Escueta, Maya, Andre Joshua Nickow, Philip Oreopoulos and Vincent Quan. 2020. "Upgrading education with technology: Insights from experimental research." *Journal of Economic Literature* 58(4):897–996.

Fiorella, Logan and Richard E Mayer. 2013. "The relative benefits of learning by teaching and teaching expectancy." *Contemporary Educational Psychology* 38(4):281–288.

Gaviria, Alejandro and Steven Raphael. 2001. "School-based peer effects and juvenile behavior." *Review of Economics and Statistics* 83(2):257–268.

Golsteyn, Bart HH, Arjan Non and Ulf Zölitz. 2021. "The impact of peer personality on academic achievement." *Journal of Political Economy* 129(4):1052–1099.

Guerrero, Tricia A and Jennifer Wiley. 2021. "Expecting to teach affects learning during study of expository texts." *Journal of Educational Psychology* 113(7):1281.

Heckman, James J. 2007. "The economics, technology, and neuroscience of human capability formation." *Proceedings of the national Academy of Sciences* 104(33):13250–13255.

Heckman, James J. 2008. "Schools, skills, and synapses." *Economic inquiry* 46(3):289–324.

Hoxby, Caroline M and Gretchen Weingarth. 2005. Taking race out of the equation: School reassignment and the structure of peer effects. Technical report Citeseer.

Imberman, Scott A, Adriana D Kugler and Bruce I Sacerdote. 2012. "Katrina's children: Evidence on the structure of peer effects from hurricane evacuees." *American Economic Review* 102(5):2048–2082.

Kolmar, Martin et al. 2013. *Group conflicts: where do we stand?* School of Economics and Political Science, Department of Economics, Univ.

Konrad, Kai A. 2009. *Strategy and dynamics in contests*. Oxford University Press.

Lavy, Victor, Olmo Silva and Felix Weinhardt. 2009. The good, the bad and the average: Evidence on the scale and nature of ability peer effects in schools. Technical report National Bureau of Economic Research.

Lefgren, Lars. 2004. "Educational peer effects and the Chicago public schools." *Journal of urban Economics* 56(2):169–191.

Muralidharan, Karthik, Abhijeet Singh and Alejandro J Ganimian. 2019. "Disrupting education? Experimental evidence on technology-aided instruction in India." *American Economic Review* 109(4):1426–1460.

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

Sacerdote, Bruce. 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*. Vol. 3 Elsevier pp. 249–277.

Shah, Rohen. N.d. "When the Student Becomes the Master: A Field Experiment on Learning by Teaching." . Forthcoming.

Sheremeta, Roman M. 2018. "Behavior in group contests: A review of experimental research." *Journal of Economic Surveys* 32(3):683–704.

Shure, Nikki. 2021. "Non-cognitive peer effects in secondary education." *Labour Economics* 73:102074.

Stinebrickner, Ralph and Todd R Stinebrickner. 2006. "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds." *Journal of public Economics* 90(8-9):1435–1454.

Tajfel, Henri, John C Turner, William G Austin and Stephen Worchel. 1979. "An integrative theory of intergroup conflict." *Organizational identity: A reader* 56(65):9780203505984–16.

Tajfel, Henri, Michael G Billig, Robert P Bundy and Claude Flament. 1971. "Social categorization and intergroup behaviour." *European journal of social psychology* 1(2):149–178.

Vigdor, Jacob and Thomas Nechyba. 2007. "Peer effects in North Carolina public schools.".

Wagener, Andreas and Martin Kolmar. 2011. "Group Identities in Conflicts.".

Xu, Di, Qing Zhang and Xuehan Zhou. 2022. "The impact of low-ability peers on cognitive and noncognitive outcomes: Random assignment evidence on the effects and operating channels." *Journal of Human Resources* 57(2):555–596.

Table 1: Student Group Details

	Team Gr.	Individual Gr.	Control Gr.	Total
Schools entering the experiment in August 2023				
No of Schools	9	8	9	9
No of All classes	16	11	13	40
No of Students	376	329	297	1002
Schools entering the experiment in February 2024				
No of Schools	4	4	4	4
No of All classes	17	16	13	46
No of Students	820	688	600	2108
All Schools				
No of Schools	13	12	13	13
No of All classes	33	27	26	86
No of Students	1196	1017	897	3110

Table 2: Baseline Test Scores Comparison

Baseline Scores	N	Control Gr.		N	Individual Gr.		N	Team Gr.	
		Mean	Std. Dev.		Mean	Std. Dev.		Mean	Std. Dev.
All									
Cognitive Score	766	0.49	1.25	926	0.49	1.18	1,118	0.43	1.28
Generosity Score	703	-2.03	0.14	861	-2.02	0.15	981	-2.03	0.18
Reciprocity Score	661	-1.94	0.23	845	-1.91	0.23	981	-1.91	0.25
Trust Score	703	-2.03	0.18	859	-2.01	0.18	981	-2.03	0.20
Noncognitive Score	661	-0.02	0.87	843	0.14	0.85	980	0.02	0.99
Low Ability									
Cognitive Score	307	-0.52	0.84	355	-0.42	0.83	450	-0.52	0.97
Generosity Score	250	-2.03	0.15	315	-2.02	0.15	374	-2.03	0.18
Reciprocity Score	233	-1.95	0.23	308	-1.90	0.22	374	-1.91	0.25
Trust Score	250	-2.02	0.19	315	-2.01	0.18	374	-2.02	0.20
Noncognitive Score	233	-0.01	0.90	308	0.16	0.83	373	0.04	0.96
Medium Ability									
Cognitive Score	238	0.69	0.81	318	0.66	0.83	354	0.63	0.90
Generosity Score	197	-2.02	0.15	271	-2.02	0.15	292	-2.03	0.19
Reciprocity Score	183	-1.94	0.24	265	-1.92	0.21	292	-1.93	0.26
Trust Score	197	-2.04	0.18	270	-2.01	0.18	292	-2.04	0.19
Noncognitive Score	183	-0.05	0.90	263	0.12	0.84	292	-0.02	1.01
High Ability									
Cognitive Score	221	1.66	0.94	253	1.57	0.94	314	1.56	0.99
Generosity Score	189	-2.03	0.13	219	-2.03	0.15	263	-2.04	0.18
Reciprocity Score	178	-1.93	0.22	217	-1.93	0.24	263	-1.90	0.25
Trust Score	189	-2.04	0.18	219	-2.01	0.18	263	-2.03	0.21
Noncognitive Score	178	0.02	0.81	217	0.08	0.90	263	0.00	1.01

Table 3: Assessment Comparisons

Variables	Individual Gr.			Team Gr.		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.
No. of HW Offered	1,016	16.60	4.28	1,156	16.30	4.43
No. of HW Attended	1,016	11.31	5.51	1,156	11.15	5.40
Average HW Score	954	0.21	0.90	1,102	0.22	0.77
No. of Tournaments Offered	1,016	3.81	1.87	1,156	3.93	1.71
No. of Tournaments Attended	1,016	2.45	1.72	1,156	2.54	1.61
Average Tournament Score	889	-0.18	1.02	1,041	0.16	0.96

Table 4: Endline Test Scores Comparison

Endline Scores	N	Control Gr.		N	Individual Gr.		N	Team Gr.	
		Mean	Std. Dev.		Mean	Std. Dev.		Mean	Std. Dev.
All									
Cognitive Score	803	0.38	1.04	901	0.46	1.05	1,089	0.53	1.10
Generosity Score	733	-2.07	0.19	808	-2.07	0.17	946	-2.09	0.19
Reciprocity Score	733	-1.98	0.26	807	-1.96	0.25	948	-2.01	0.25
Trust Score	733	-2.07	0.20	803	-2.08	0.19	948	-2.11	0.19
Noncognitive Score	733	0.05	1.04	801	0.08	0.92	945	-0.11	1.03
Low Ability									
Cognitive Score	278	0.19	1.02	321	0.25	1.01	409	0.32	1.02
Generosity Score	251	-2.06	0.18	292	-2.06	0.16	345	-2.07	0.19
Reciprocity Score	251	-1.98	0.27	292	-1.93	0.26	346	-1.99	0.25
Trust Score	251	-2.07	0.21	290	-2.07	0.19	346	-2.09	0.19
Noncognitive Score	251	0.04	1.06	290	0.20	0.88	345	0.00	1.01
Medium Ability									
Cognitive Score	220	0.34	0.98	284	0.48	1.02	328	0.59	1.13
Generosity Score	198	-2.07	0.19	250	-2.07	0.17	291	-2.10	0.19
Reciprocity Score	198	-1.98	0.26	249	-1.96	0.24	291	-2.03	0.26
Trust Score	198	-2.09	0.20	249	-2.08	0.18	291	-2.12	0.18
Noncognitive Score	198	0.03	1.05	248	0.08	0.91	290	-0.16	1.06
High Ability									
Cognitive Score	205	0.67	1.07	233	0.74	1.05	288	0.81	1.11
Generosity Score	189	-2.07	0.18	209	-2.10	0.16	251	-2.12	0.18
Reciprocity Score	189	-1.99	0.26	209	-2.00	0.25	252	-2.03	0.24
Trust Score	189	-2.07	0.19	207	-2.10	0.19	252	-2.11	0.19
Noncognitive Score	189	0.06	0.99	206	-0.06	0.93	251	-0.23	1.04

Table 5: Balance Check for Cognitive Scores

	(1)	(2)	(3)	(4)	(5)
	All(Base)	All(Score)	LA	MA	HA
Team Gr	0.01	-0.01	0.02	-0.02	0.02
	(0.14)	(0.15)	(0.16)	(0.16)	(0.15)
Individual Gr	0.13	0.13	0.16	0.10	0.12
	(0.14)	(0.14)	(0.14)	(0.15)	(0.16)
Constant	0.42***	0.44***	-0.55***	0.63***	1.55***
	(0.10)	(0.10)	(0.11)	(0.11)	(0.11)
Observations	2810	2566	1008	832	726
R^2	0.233	0.225	0.333	0.525	0.508
School FE	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
All(Base) - All Students with Baseline Scores
All(Score) - All Students with Baseline and Endline Scores
HA=High Ability, MA=Medium Ability, LA=Low Ability
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 6: Balance Check for Noncognitive Scores

	(1) All Basline	(2) All	(3) LA	(4) MA	(5) HA
Team Gr	-0.00	-0.03	-0.07	0.03	-0.05
	(0.12)	(0.12)	(0.13)	(0.13)	(0.14)
Individual Gr	0.09	0.07	0.05	0.14	0.02
	(0.10)	(0.11)	(0.13)	(0.12)	(0.13)
Constant	0.02	0.02	0.07	-0.06	0.03
	(0.09)	(0.09)	(0.10)	(0.09)	(0.10)
Observations	2484	2061	754	609	556
R^2	0.114	0.111	0.114	0.141	0.140
School FE	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
All Baseline - All Students with Baseline Scores
All Score - All Students with Baseline and Endline Scores
HA=High Ability, MA=Medium Ability, LA=Low Ability
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 7: Class Level Cognitive Score Estimates

	(1) Avg Score	(2) COV	(3) High Avg. Score	(4) Med Avg. Score	(5) Low Avg. Score
Team Gr.	0.23**	-0.03*	0.22*	0.33**	0.21
	(0.11)	(0.02)	(0.13)	(0.13)	(0.13)
Individual Gr.	0.10	-0.00	0.05	0.22	0.08
	(0.11)	(0.02)	(0.13)	(0.14)	(0.13)
Avg Class Baseline Score	0.17**				
	(0.08)				
COV Class Baseline Score		0.08			
		(0.06)			
HA Class Avg Baseline Score			0.24**		
			(0.09)		
MA Class Avg Baseline Score				0.26***	
				(0.10)	
LA Class Avg Baseline Score					0.16*
					(0.09)
Constant	0.27***	0.30***	0.29*	0.10	0.22*
	(0.08)	(0.03)	(0.16)	(0.11)	(0.11)
Observations	86	85	84	85	86
R^2	0.720	0.474	0.694	0.665	0.586
School FE	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
HA= High Ability, MA= Medium Ability, LA= Low Ability
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 8: Student Level Skill Production Estimates

	(1) All	(2) HA	(3) MA	(4) LA	(5) All	(6) HA	(7) MA	(8) LA
Team Gr.	0.25***	0.28**	0.28**	0.20*	0.17*	0.20*	0.19*	0.15
	(0.09)	(0.12)	(0.11)	(0.11)	(0.10)	(0.11)	(0.11)	(0.14)
Individual Gr.	0.07	0.09	0.11	0.03				
	(0.09)	(0.13)	(0.11)	(0.12)				
Baseline Score	0.25***	0.36***	0.34***	0.22***	0.25***	0.33***	0.32***	0.23**
	(0.03)	(0.07)	(0.07)	(0.08)	(0.03)	(0.07)	(0.09)	(0.10)
Constant	0.23***	0.03	0.12	0.28***	0.30***	0.15	0.23**	0.32***
	(0.06)	(0.12)	(0.10)	(0.08)	(0.07)	(0.15)	(0.09)	(0.11)
Observations	2566	726	832	1008	1863	521	612	730
R^2	0.235	0.262	0.260	0.171	0.238	0.292	0.244	0.160
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
HA=High Ability, MA=Medium Ability, LA=Low Ability
All scores normalized to μ=0 and σ=1
Estimates in columns(1)-(4) are on full data
Estimates in columns(5)-(8) are for Team and Ind classes
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 9: Effect of Self and Team Feedback on Skill Production

	(1) IV(All)	(2) IV(HA)	(3) IV(MA)	(4) IV(LA)	(5) OLS(All)	(6) OLS(Team+Ind)
Team Feedback	0.11***	0.12**	0.12**	0.09*	0.10**	0.12**
	(0.04)	(0.06)	(0.05)	(0.05)	(0.04)	(0.06)
Self Feedback	0.11	0.13	0.16	0.05	0.16	0.21
	(0.13)	(0.17)	(0.16)	(0.17)	(0.12)	(0.19)
Baseline Score	0.24***	0.36***	0.33***	0.21***	0.24***	0.25***
	(0.03)	(0.07)	(0.07)	(0.07)	(0.03)	(0.03)
Constant					0.23***	0.18
					(0.05)	(0.12)
Observations	2566	726	832	1008	2566	1863
R^2	0.087	0.082	0.059	0.034	0.236	0.241
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
IV(ALL)-IV model for full Sample
IV(HA/MA/LA)-IV model for High, Medium, and Low Ability students samples
OLS(ALL/Team+Ind)-OLS model for all and only Team and Individual students
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 10: Effect of Team Interactions on Skill Production

	(1) IV(All)	(2) IV(HA)	(3) IV(MA)	(4) IV(LA)	(5) OLS(ALL)	(6) OLS(Team+Ind)
Avg. Team Help	0.24***	0.26**	0.27**	0.20*	0.21***	0.24**
	(0.09)	(0.12)	(0.11)	(0.11)	(0.07)	(0.11)
Avg. Self-Help	0.11	0.13	0.17	0.05	0.15	0.18
	(0.13)	(0.17)	(0.16)	(0.17)	(0.12)	(0.19)
Baseline Score	0.24***	0.36***	0.33***	0.21***	0.24***	0.24***
	(0.03)	(0.07)	(0.07)	(0.07)	(0.03)	(0.03)
Constant					0.23***	0.20*
					(0.05)	(0.12)
Observations	2566	726	832	1008	2566	1863
R^2	0.086	0.078	0.059	0.035	0.236	0.240
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
IV(ALL)-IV model for full sample
IV(HA/MA/LA)-IV model for high, medium,and low ability student samples
OLS(ALL/Team+Ind)-OLS model for all and only team and individual students
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 11: Effect of Teammates Homework Performance

	(1) Tourn Score(OLS)	(2) Tourn Score(IV)	(3) Endline Score(OLS)	(4) Endline Score(IV)
Teammate Avg. Topic HW Score	0.14**	0.10*	0.08**	0.08**
	(0.05)	(0.06)	(0.03)	(0.04)
Classmate Avg. Topic HW Score	0.24***	0.24***	0.12***	0.12***
	(0.05)	(0.05)	(0.03)	(0.03)
Baseline Score	0.10***	0.10***	0.12***	0.12***
	(0.03)	(0.03)	(0.03)	(0.03)
Constant	-0.01		-0.06	
	(0.04)		(0.04)	
Observations	3420	3420	6922	6922
R^2	0.164	0.070	0.066	0.036
School FE	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes
Topic FE	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
All scores normalized to μ=0 and σ=1
Tourn Score(OLS/IV) = OLS/IV estimates on student's topic-wise torunament scores
Endline Score(OLS/IV) = OLS/IV estimates on student's topic-wise endline scores
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 12: Relationship between tournament rank and baseline cognitive scores

	(1) Ind	(2) Team	(3) TeamHR+Ind +Ctrl	(4) TeamHR+Ind
Baseline Score	0.07***	0.00	-0.02	-0.03
	(0.02)	(0.01)	(0.02)	(0.02)
Constant	0.48***	0.46***	0.24***	0.36***
	(0.01)	(0.01)	(0.04)	(0.06)
Observations	926	1118	2246	1480
R^2	0.025	0.034	0.075	0.129
School FE	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
All scores normalized to μ=0 and σ=1
Ind-All Individual students
Team-All Team students
TeamHR+Ind+Ctrl-High tournament rank Team students+(Individual+Control)students
TeamHR+Ind-High tournament rank Team students +Individual students
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 13: Effect of Tournament Rank on Skill Production

	(1) OLS(All)	(2) OLS(Team+Ind.)	(3) IV(All)	(4) IV(Team+Ind)
High Team Rank	0.26***	0.23**	0.52***	0.36*
	(0.09)	(0.10)	(0.19)	(0.21)
Low Team Rank	0.11	0.07		
	(0.10)	(0.11)		
Baseline Score	0.25***	0.25***	0.24***	0.25***
	(0.03)	(0.03)	(0.03)	(0.03)
Constant	0.28***	0.32***		
	(0.05)	(0.07)		
Observations	2566	1863	2566	1863
R^2	0.235	0.240	0.077	0.077
School FE	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
OLS(All/Team+Ind)-OLS model for all and only team and individual students
IV(All/Team+Ind)-IV model for all and only team and individual students
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 14: Effect of Tournament Rank on Skill Production Across Students of Different Abilities

	(1) OLS(HA)	(2) OLS(MA)	(3) OLS(LA)	(4) IV(HA)	(5) IV(MA)	(6) IV(LA)
High Team Rank	0.28**	0.24*	0.19	0.40*	0.39*	0.33
	(0.12)	(0.13)	(0.14)	(0.23)	(0.21)	(0.31)
Low Team Rank	0.04	0.08	0.10			
	(0.14)	(0.11)	(0.16)			
Baseline Score	0.35***	0.32***	0.23**	0.36***	0.32***	0.23**
	(0.07)	(0.09)	(0.10)	(0.08)	(0.09)	(0.10)
Constant	0.15	0.25***	0.33***			
	(0.15)	(0.09)	(0.11)			
Observations	521	612	730	521	612	730
R^2	0.296	0.245	0.161	0.064	0.038	0.027
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
OLS(HA/MA/LA)-OLS model for the sample of high/medium/low ability team and individual students
IV(HA/MA/LA)-IV model model for the sample of high/medium/low ability team and individual students
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 15: Student Level Non-Cognitive Skill Production Estimates

	(1) All	(2) High Ability	(3) Medium Ability	(4) Low Ability
Team Gr. Peer	-0.19	-0.25	-0.24	-0.13
	(0.12)	(0.16)	(0.15)	(0.14)
Team Gr. Not Peer	-0.24**	-0.34**	-0.23	-0.18
	(0.12)	(0.14)	(0.18)	(0.15)
Individual Gr.	-0.09	-0.24*	-0.02	-0.01
	(0.09)	(0.13)	(0.12)	(0.10)
Baseline NonCog Score	0.10***	0.13**	0.07	0.09*
	(0.04)	(0.06)	(0.06)	(0.05)
Constant	0.09	0.08	0.03	0.14
	(0.07)	(0.08)	(0.11)	(0.09)
Observations	2065	557	611	755
R^2	0.101	0.111	0.116	0.115
School FE	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
* $p < .1$, ** $p < .05$, *** $p < .01$

Appendix: Improving Skill Production with Peer-Driven Knowledge Sharing in K-12 Schools

by Anuj Kumar and Anurag Garg

Appendix

A EPInc platform

The EPInc platform is designed to improve skill production in team classes with the following mechanisms.

A.1 Feedback

The EPInc platform provides feedback to team members on their knowledge gaps. After the end time for base homework is over, students receive notifications with topic-wise performance of their team members. An example of such a notification is given in Figure A1, which gives scores of five team members on two math topics, "Building with Blocks" and "A Trip to Bhopal". One student (ID Zikra_2347) has her homework performance displayed in red color because s/he was absent from the homework. Thus, this screen provides student feedback on the topic-wise performance of her teammates who have attended the homework. Students can utilize this feedback to seek help on their weak topics from the relevant teammate and offer help to teammates on topics they have done well.

[Figure A1 about here.]

Teachers/volunteers were trained to keep reminding students to check the topic-wise performance of their team members and discuss those topics with each other. At the end of each base homework period, volunteers were trained to ask members (of a few randomly selected teams) to open the topic-wise performance feedback screen on their computer tablets and explain how they should utilize this information to share knowledge among themselves.

A.2 Improve learning motivation

The team tournaments on the EPInc platform may motivate students to invest in studies to help their team win tournaments. The Platform is carefully designed to create a feeling of team spirit among students by allowing them to choose their own team names and mascots. The student's home screen on the platform prominently displays the team name and mascot at the top. For example, on logging on to the platform, the student sees her team name, "Team UF," and team mascot, "Longhorn" in Figure A2. This is designed to remind students of their team identity every time they log on to the platform.

[Figure A2 about here.]

The student's home screen further shows the performance of the student's team vis-a-vis the top-performing team in the homework. For example, the student's team, "Team UF," is in second place with a 50% score, and the first-ranked team, "Team RWF," has scored 90%. This design could motivate team members to catch up with the top team. In addition, a student is shown the performance of her teammates at the bottom of the screen, which she can utilize to motivate and support teammates who are lagging behind by giving them emojis and claps. For example, the student motivates the teammate with a 20% score by giving her an emoji in Figure A3.

[Figure A3 about here.]

After the tournament completion, teachers display the names and mascots of three top-ranked teams on the social leader board in the class. Figure A4 shows the class leaderboard used in our experiment. The social recognition of being part of the displayed winning team in class should motivate students to contribute to their team performance.

[Figure A4 about here.]

A.3 Induce knowledge sharing

Knowing each others' knowledge gaps and motivation to win tournaments may induce students to share knowledge with their team members. Toward this objective, the EPInc platform reminds students to interact with their team members on relevant screens.

When students start the personalized homework, the first screen reminds them whether they consulted teammates on their weak topics, indicating their weak topics and the teammates who could help. The last screen of the personalized homework asks students to name the team members with whom they have interacted/discussed their weak topics. These screen designs deliberately remind students to discuss math topics with their team members.

The EPInc platform notifies students about upcoming tournaments and urges them to prepare with their teammates. For example, students get notified about the date, time, and topics covered in the upcoming tournament, as shown in Figure A5. The students are urged to discuss these topics to do well in the tournament.

[Figure A5 about here.]

B Relationship Between Self and Team Feedback with Cognitive Scores

High-ability students, on average, could be more engaged with education and, thus, attend a higher proportion of offered homework. Thus, we expect the proportion of homework attended by students in individual classes to be positively correlated to their baseline scores. However, the relationship between the average number of team members attending homework with the student's cognitive abilities is unclear. We examine these facts in our data with the following specifications:

$$Y_i = \beta_1 BaselineScore_i + \alpha_s + \alpha_g + \epsilon_i \tag{9}$$

where Y_i is (1) the proportion of total offered homework attended by student i in individual classes (self-feedback) and (2) the average number of team members (including the focal student) attending offered homework with student i in team classes (team feedback). We estimate Equation 9 for the data on individual and team classes separately for the abovementioned two dependent variables.

Table A1 reports the estimated coefficients. We find a positive and significant coefficient for $BaselineScore$ in Column (1), confirming a positive correlation between students' cognitive abilities and propensities to attend homework. In column (2), we find a positive but insignificant estimate for Team feedback, suggesting that the number of teammates attending homework with the student is uncorrelated to her cognitive abilities.

[Table A1 about here.]

C Non-cognitive skills measurement

This section discusses the social preference experiments to measure students' non-cognitive skills of generosity, reciprocity, and trust. We conducted these experiments for the three categories of classes – team, individual, and control classes.

Instructions – Teachers provide the following instructions to students before starting the experiment.

- You are given money to play three games. In each game, you will be asked to decide how much money you would give to your classmate whose name appears on the top right of the paper slip given to you.
- For each game, carefully listen to the teacher's instructions and make your decisions individually without talking to or showing your response to others.
- Provide your response on the provided paper slips. Your responses are private and will not be shared with others.
- After sharing money with your classmates, you can buy candies and school supplies from the supply store outside your classroom with the remaining money.
- After the games, please fold your slip and give it to the teacher.

[Figure A6 about here.]

Dictator Game-1 – We give you Rs 10 to play this game. You can give any amount between Rs. 0 and Rs. 10 to your classmate whose name appears on the slip. How much money will you give to your classmate? Provide your response individually in the provided slip.

Reciprocity Game-2 – We give Rs 10 each to you and your classmate whose name appears on the slip to play this gift exchange game. Your classmate has given you a gift of Rs. 5 from his/her Rs. 10. Now you have Rs. 15. You can return any amount between 0 and 15. How much money will you give to your classmate? Provide your response individually in the provided slip.

Trust Game-3 – We give Rs 10 each to you and your classmate whose name appears on the slip to play this game. You can give your classmate any amount between Rs. 0 and Rs. 10. We will triple that amount and give it to your classmate. Your classmate can return any amount between Rs. 0 and the total amount he has. However, he is not required to return anything. If you share more with your classmate, she/he will get a bigger amount (three times what you shared). If she/he shares back with you, you both will be better off. However, you will be worse off if she/he doesn't share back. How much money (Rs.) will you give to your classmate? Provide your response individually in the provided slip.

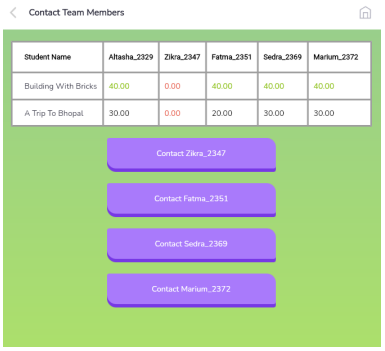


Figure A1: Topic-wise Team Homework Performance Screen

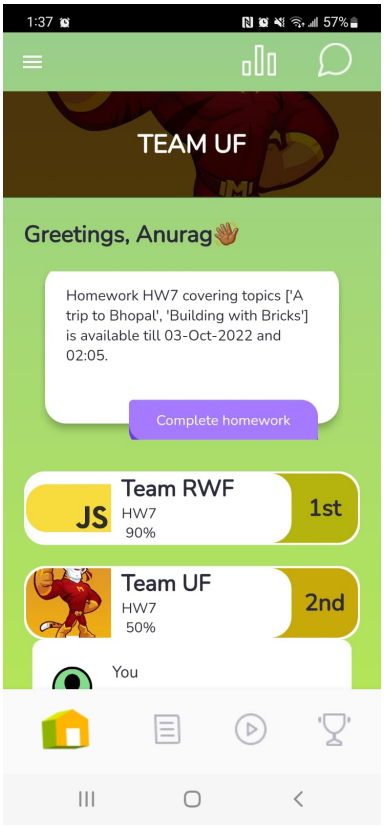


Figure A2: Student's Home screen on EPInc Platform

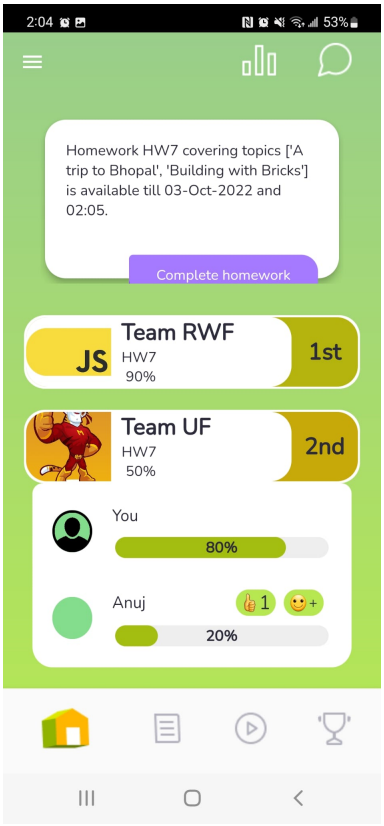


Figure A3: Student's Home screen on EPInc Platform Showing Team Member Performance



Figure A4: Leaderboard displaying Top Three Ranked Teams in a Class

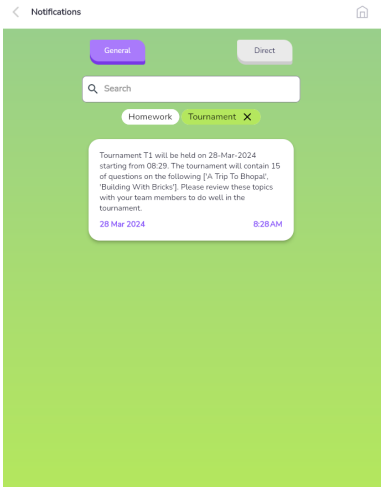


Figure A5: Tournament Notification to Students

Student Name -

Classmate Name -

Class –

School -

	Game 1	Game 2	Game 3
Money with you (Rs.)	10	10 + 5 = 15	10
How much will you contribute to (share with) your classmate			

Figure A6: Non-cognitive Scores Slip

Table A1: Relationship of self and team feedback with students' cognitive ability

	(1) SF(Ind)	(2) TF(Team)
Baseline Score	0.02**	0.03
	(0.01)	(0.03)
Constant	0.70***	2.19***
	(0.01)	(0.06)
Observations	926	1118
R^2	0.073	0.203
School FE	Yes	Yes
Grade FE	Yes	Yes

Standard errors in parentheses
Standard errors cluster corrected at the class level
All scores normalized to μ=0 and σ=1
SF(Ind) = Self Feedback in Individual Classes.
TF(Team) = Team Feedback in Team Classes
* $p < .1$, ** $p < .05$, *** $p < .01$