Improving Educational Technology with Collaborative Design

Mohsen Foroughifar
Carnegie Mellon University

Anuj Kumar
University of Florida

Abstract

Most educational technology (EdTech) platforms personalize learning but overlook classmates as a learning resource. We examine whether and why putting students in teams improves EdTech effectiveness by building an education platform and conducting a randomized field experiment across 92 grade 3–6 math classes in 13 Indian schools. Classes were randomly assigned to either a no platform condition, an individual platform condition with practice assignments and individual class tournaments, or a team platform condition with students collaborating in ability-balanced teams for assignments and tournaments while the platform displays teammates' knowledge gaps in assignments. We find that students in the team condition help teammates shown as lower-performing, and collaboration between teammates is associated with higher tournament scores. The team condition increases end-of-year math scores by 3.8 points (0.23 SD) relative to control; the individual design's effect is positive but has a lot of variation. To disentangle the contributions of platform technology, teams, and their interaction, we develop a structural model of students' assignment practice, teammate assistance, and tournament attendance. Counterfactual analyses attribute approximately 31% of the team condition's gain over control to the platform technology alone, 58% to teams alone, and 11% to their interaction through help enabled by platform information about teammates. Without this information, teammate assistance would fall by approximately 17%. The team design increases practice most among the least productive students and improves achievement across all levels of learning productivity. These results highlight the role of organizational design in educational technology and suggest that collaboration can strengthen the returns to policies that encourage participation.

Keywords. skill production, K–12 schools, peer effects, group contests, technology and organization, structural models

Acknowledgments. We thank Samina Bano, Shipra Srivastava, Prakhar Dwivedi, and Anirban Samanta of the RightWalk Foundation (RWF), for their extremely generous contribution of financial and manpower resources to conduct the experiment in multiple schools. We also thank K S Binu of MOBIUX (mobiux.com), the software firm that developed and maintained the software platform during the experiment. Without this support, this study would not have been possible.

1 Introduction

Educational technology (EdTech) platforms have expanded access to personalized practice and scalable instruction (Escueta et al., 2020; Bulman and Fairlie, 2016), yet most treat students as isolated learners and overlook their classmates as a well-documented source of learning. How students are grouped and whether they interact productively can shape what they learn (Sacerdote, 2001; Epple and Romano, 2011), but putting students together without the right incentives and information may not always help (Carrell et al., 2013), and collaborative activities that work in classrooms may not automatically carry over to a digital environment. A central question for EdTech design is, therefore, whether a platform can be made more effective by organizing interactions so that each student's practice also helps classmates learn.

Past research has shown that returns to information technology can depend on complementary organizational practices (Bresnahan et al., 2002; Bloom et al., 2012). An analogous possibility arises in schools, where educational technology can be more valuable when it provides digital feedback in a setting where classmates can collaboratively learn in teams. However, measuring the joint effects of technology and team collaboration is challenging because students often select into specific teams in class, and introducing a platform while changing the structure of class activities to a team setting combines different effects together. Evaluating such complementarities experimentally requires variations in the relevant components separately and sufficient power to estimate their interactions (Muralidharan et al., 2025).

We study how a collaborative team design changes the effectiveness of an education platform by building and deploying an EdTech platform in kindergarten-to-grade-12 (K–12) math classes. Building the software ourselves, let us implement the desired experimental designs and record student engagement throughout. The core functionality of the platform is to diagnose students' knowledge gaps from practice assignments, assign personalized remediation, and run recurring cycles of practice and in-class tournaments, with classroom recognition for top performers. To examine the effectiveness of the platform in different conditions, we create two designs of the platform. In its individual design, feedback and recognition are student-specific. Each student sees only their own knowledge gaps, and the top scorers in each tournament are recognized individually. In the team design, we randomly create ability-balanced teams so that students can work in teams while observing their teammates' knowledge gaps, and the platform recognizes the team's performance in tournaments, even though students still answer tournament questions individually. Together, these features give students both the information and the incentive to help one another, and can turn

classmates from passive co-users of the platform into contributors to one another's learning.

We ask four research questions: 1) What is the total effect of team design of an EdTech platform on students' math achievement? 2) How much of the total effect comes from the technology itself, collaborations within teams, and the interaction of the two (i.e., the help that the platform's information about teammates makes possible)? 3) Which students benefit the most and through what mechanism? 4) How should EdTech platforms take advantage of the team design to improve learning?

To answer these questions, we partnered with a non-governmental organization[1] and ran a randomized field experiment on our platform over two academic years in 92 grade 3–6 math classes in 13 Indian K–12 schools. The schools are located in 6 different cities across 3 states and have altogether 3,294 students. Within each school, classes were randomly assigned to one of three conditions: i) a control condition in which the class continued with paper assignments and no platform, ii) an individual platform condition (arms), where students practice and compete in tournaments on their own, or iii) a team platform condition, where students are randomly grouped into ability-balanced teams, see one another's knowledge gaps on the platform, and compete as teams in tournaments. The two platform conditions follow the same curriculum as the control and, throughout the school year, repeat cycles of the same nature on the platform in which students participate in a diagnostic assignment, a personalized assignment and a class tournament.[2]

Our experimental intent-to-treat effect show that students assigned to team classes scored 3.8 points (0.23 SD) higher than control students on the end-of-year assessment ($p < 0.05$). Students in the individual-design classes also scored higher than in control classes, but the difference is small with a high variance. The difference between the team and individual arms is 2.4 points, but not statistically significant at conventional levels. The average practice and tournament attendance are similar across the platform arms. However, within teams, students give more help when more teammates' assignment scores fall below the team median, conditional on how many teammates are present. Collaboration is also positively associated with tournament performance within students, conditional on practice, and average tournament performance predicts end-of-year achievement. These patterns suggest a collaboration mechanism, although helping and diagnostic performance are not themselves randomized.

The experiment does not separately identify the contributions of teams and of sharing the platform's diagnostic reports among teammates compared to the effect of technology alone. It also leaves the distributional effects across students opaque. Two traits that matter espe-

[1]The RightWalk Foundation (rightwalk.org) helps economically disadvantaged children enroll in private schools under India's Right to Education Act.

[2]The experiment is pre-registered on ASPredicted (#97643, `https://aspredicted.org/see_one.php`); Section 3 describes how the final sample differs from the registered one.

cially for students' practice and tournament participation choices are learning productivity (learning per unit of practice) and the cost of practice (cost per unit of practice). Both of these traits are latent in the data, so the reduced form cannot recover their distributions across students or how behavior varies with them. We therefore develop a structural model of assignment practice, helping within teams, tournament attendance, and skill accumulation to disentangle these effects and to examine counterfactual scenarios beyond the experiment.

Our model reveals that one collaboration within a team produces 0.44 end-of-year points of knowledge. Compared to control, technology itself accounts for 31% of the achievement gain in the team condition; the addition of teams without sharing the platform's diagnostic reports among teammates accounts for 58% of the gain; the interaction of technology and teams, i.e., sharing these reports among teammates, accounts for 11% of the gain. Not sharing the reports among teammates reduces predicted collaboration by about 17%. This decomposition is conditional on the platform's other functions remaining available and does not identify the effect of teams without technology.

With the estimated model, we also evaluate the distribution of gains and the consequences of changing particular design features through counterfactual simulations. First, we compare each team-class student under the team design and under the individual design. Practice rises most among students with low learning productivity. However, end-of-year achievement rises for 99.5% of students, by 2.4 points on average and by similar amounts at every level of learning productivity, because a collaboration produces the same knowledge across productivity levels. Second, a nudge that increases students' assignment attendance can increase achievement by 0.28 in individual classes versus 0.35 points in team classes. The 28% higher achievement in the latter case is because in a team class a student brought to work gains both from their own practice and from the help that work attracts; 60% of the difference accrues to the less productive half of students.

In summary, our study provides three insights. First, personalized learning and collaboration can serve complementary roles in classroom technology. Second, a teammate's help directed toward students with lower performance can increase the return to practice and improve end-of-year math scores. Third, while students with lower learning productivity are more likely to increase practice in teams (versus individually), all students benefit from collaboration in teams. These results inform the current debate over replacing classrooms (at least partially) with hyper-personalized content generated for individual students using advanced AI and machine learning tools.[3] Our design couples balanced teams with tools that make teammates' knowledge gaps visible, which encourages collaboration among team members. Once that environment is in place, managers and evaluators may benefit more from

[3] `https://alpha.school/`

higher returns to policies that encourage more students to attempt practice assignments.

2 Related Literature and Contributions

Our study connects literature in economics and information systems on educational technology, collaborative learning, and organizational complementarities.

Personalized educational technology diagnoses learning gaps and adapts content to students' demonstrated skills. This approach builds on evidence that instruction matched to students' learning levels can improve achievement (Banerjee et al., 2007, 2016). Specifically, experimental studies in India document substantial gains from personalized adaptive learning (PAL) (Muralidharan et al., 2019; Muralidharan and Singh, 2025). However, literature reviews emphasize that effects can vary with the technology and its implementation (Escueta et al., 2020; Bulman and Fairlie, 2016). Specifically, PAL applications help students who are willing to engage with adaptive content and can learn from it independently. So, most successful personalized learning interventions have ensured that students obtain the necessary support with PAL by offering it during or after school hours under teachers' supervision (Muralidharan et al., 2019; Banerjee et al., 2007). We contribute to this literature by showing that classroom peers, if appropriately incentivized, can serve as an alternative resource to help students engage with and learn from technology-mediated personalized learning. The contrast of interest therefore concerns the organization of learning around a common technology.

The literature on collaborative learning has established that classmates and other peers can influence achievement, but the consequences of changing peer composition depend on the setting and the interactions it generates (Sacerdote, 2001; Carrell et al., 2009; Epple and Romano, 2011; Sacerdote, 2011). Some studies document that students gain the most when surrounded by peers of similar ability, and mixing student with different abilities may be counter-productive (Sacerdote, 2011; Hoxby and Weingarth, 2005; Lavy et al., 2009). However, others have found that mixing students with different abilities in a class is beneficial (Betts and Shkolnik, 2000; Lefgren, 2004), and low-ability students benefit the most from high-ability peers (Vigdor and Nechyba, 2007; Burke and Sass, 2013). Thus, the impact of mixing students of different abilities in a class on educational outcomes is unclear. Moreover, randomly assigning students to groups does not necessarily produce the intended social interactions (Carrell et al., 2013).

Our platform addresses this problem by combining ability-balanced teams with diagnostic reports on learning gaps shared among teammates, collective recognition of team members in tournaments, and tracking help among teammates directly, which allows us to relate help

among teammates to the information available within teams and students' performances over the academic year. This design is inspired by experiments that examine how financial incentives, recognition, and behavioral framing affect student behavior (Fryer, 2011; Levitt et al., 2016; Cotton et al., 2026).

Our platform design also draws on information systems research on how learning technology design shapes outcomes (Alavi and Leidner, 2001; Gupta and Bostrom, 2009). Gamification and contest-based recognition can be used to sustain engagement, yet their effects depend on how they are designed and implemented (Liu et al., 2017; Hamari et al., 2014). Digital nudges, small changes to a platform's choice environment, can also move user behavior at scale (Weinmann et al., 2016) with effects hinging on how tools are configured and combined (Escueta et al., 2020; Bulman and Fairlie, 2016; Kumar and Garg, 2025). We contribute to this literature by providing evidence on the combined benefits of sharing diagnostic reports among teammates, collaboration, and team contests, together with a model that decomposes the contribution of these features.

Finally, our analysis relates to the literature on complementarity between technology and organization. Studies of firms show that organizational practices influence the returns to information technology (Bresnahan et al., 2002; Bloom et al., 2012). In education, evaluations often study bundles of technology and implementation practices (Escueta et al., 2020; Bulman and Fairlie, 2016). Separating the components of a bundle requires either a factorial design, whose interaction arms are costly and often underpowered (Muralidharan et al., 2025), or estimating a structured model that recovers the contribution of different elements from behavior observed inside the experiment. We follow the second route, in the tradition of work that uses randomized experiments to discipline and validate structural models (Todd and Wolpin, 2006; Attanasio et al., 2012; Duflo et al., 2012), and apply it to the interaction between an EdTech platform and the organization of students into teams.

3 Data and Experimental Design

3.1 Experimental setup

Our field experiment runs on an EdTech platform that we designed and built for this research. Building it ourselves allows us to implement the individual and team designs exactly and record practice assignments, tournaments, and recognition as research data. The educational content on the platform comes from a large bank of standardized math questions aligned with India's grade 3–6 syllabus from The National Council of Educational Research and

Training (NCERT).[4] Students use school tablets during school hours, and teachers continue their usual instruction, so the platform supplements practice and tournaments but does not replace lectures. Over the course of the experiment, the platform delivered more than 34 thousand practice assignments to students, drawn from 877 distinct assignments, and recorded more than 0.4 million item responses to questions attempted by students.

The experiment covers 92 grade 3–6 math classes in 13 schools in 6 cities and 3 Indian states over two academic years. Within schools, classes were randomly assigned to one of three conditions: a control condition with no platform (27 classes), an individual platform condition (30 classes), or a team platform condition (35 classes).[5] The three arms followed the same curriculum and retained regular classroom instruction. Practice delivery, feedback, and tournaments differed by arm.[6] Students were scheduled to complete standardized math assessments at the start and end of the school year; these scores are the baseline and endline (end-of-year) skill measures.[7]

During the school year, the two platform arms repeat the same cycle alongside regular instruction, for an average of 3.4 cycles per student. Each cycle has five steps, in which (1) teachers teach a set of math concepts and generate a diagnostic assignment from the platform's question bank; (2) students complete the diagnostic assignment on the platform; (3) the platform gives students arm-specific feedback on their knowledge gaps; (4) students work on a personalized practice assignment that the platform generates from their diagnostic results; and (5) students take part in an in-class tournament with arm-specific recognition. All steps take place in class during school hours. Control classes received regular instruction and completed a similar number of paper-based practice assignments, without feedback, personalization, or tournaments.

Relative to the individual design, the team design adds three features, namely ability-balanced teams of four students, fixed for the year and formed by drawing one student from each quartile of the class's baseline scores; shared visibility of knowledge gaps within

[4]NCERT is the federal body for K–12 education in India.

[5]The final sample differs in size from the registered experiment (ASPredicted #97643, `https://aspredicted.org/see_one.php`). We registered the experiment in 2022 for about 1,000 students and started it in August 2023 with 978 students in 40 grade 3–6 math classes in nine schools in four cities, over the academic year from July 2023 to April 2024. In January 2024, we had the opportunity to add 2,316 students to our experiment in 52 grade 3–6 math classes in four schools in two cities, over the academic year from January to December 2024. The 3,294 students are the sum of the two samples. We added the second sample to make the findings more general, but did not file an addendum to the registration. The analysis is consistent with the registered objective.

[6]The experimental procedure was approved by a university institutional review board after several rounds of review.

[7]Question papers for the baseline and endline assessments were prepared by experts outside of the schools in our experiment and consisted of standardized multiple-choice questions based on the math syllabus of the appropriate grade level as prescribed by the NCERT in India.

teams; and team recognition. Stratifying team formation by baseline achievement reduces differences in initial ability across teams, although it does not eliminate them. Team students also see progress cues for their teammates and receive reminders that encourage them to help before tournaments. In both platform designs, students attend a tournament and answer the questions individually within a fixed time. After the tournament, the class applauds the winners, and their names are posted on the classroom "Champion Board." Individual classes recognize the five attending students with the highest scores; team classes recognize the three teams with the highest sum of attending members' scores. Tied students are recognized jointly.

3.2 Sample and descriptive statistics

Table 1 reports the distribution of classes, grades, baseline and endline scores, and gender across the three arms. Because randomization assigned classes within schools, the tests of equal class counts and grade composition are χ^2 tests on the 92 classes, and the tests for student-level variables are regressions on the arm indicators with school fixed effects and standard errors clustered by class. The pooled tests do not reject equality of mean class sizes or baseline scores. Three features deserve discussion here: the grade composition, where the test on all grades does not reject ($p = 0.103$) but 14 of the 22 grade-6 classes are in the team arm (randomization-inference $p = 0.007$); the female share, which differs across arms because 44 of the 92 classes are single-sex; and the share of students without a baseline score. To measure the intent-to-treat effects, we include these factors as covariates in our regressions. Appendix A reports the full balance checks.

We also collected caste, parents' education, and family income for about 70% of students. Among respondents, approximately 92% are in the general caste category and 99% report monthly family income below INR 25,000. Approximately 49% of fathers and 46% of mothers have less than an eighth-grade education.

Of the 3,294 randomized students, 309 lack a baseline score and 373 lack an endline score. Baseline scores are missing for 14.3% of students in the control condition, 8.6% in the individual condition, and 6.4% in the team condition. Endline scores are missing for 10.1% of students in the control condition, 14.9% in the individual condition, and 9.0% in the team condition. Because 75 students lack both assessments, 607 students lack at least one assessment and 2,687 have both scores observed.

We test whether assessment availability differs between arms by regressing indicators for missing baseline and endline scores on the two treatment indicators and school fixed effects, with standard errors clustered by class. Joint tests reject equal baseline availability across

Table 1. Sample composition, cognitive scores, and gender by experimental arm

	Control	Individual	Team	All	p
A. Units of observation					
Classes	27	30	35	92	0.587
Schools	13	12	13	13	
Students	936	1,122	1,236	3,294	
Mean class size	34.7	37.4	35.3	35.8	0.855
B. Classes by grade					
Grade 3	8	8	8	24	
Grade 4	8	7	9	24	
Grade 5	8	10	4	22	
Grade 6	3	5	14	22	0.103
C. Cognitive scores					
Baseline score	42.06	41.90	41.34	41.73	0.913
	(18.69)	(17.62)	(19.30)	(18.57)	
Baseline missing (%)	14.3	8.6	6.4	9.4	0.029
Endline score	41.00	42.01	43.09	42.13	0.302
	(15.56)	(15.88)	(16.56)	(16.07)	
Endline missing (%)	10.1	14.9	9.0	11.3	0.425
Gain, endline − baseline	−1.33	−0.08	1.83	0.33	0.332
D. Gender and estimation sample					
Female (%)	31.2	57.3	51.1	47.6	0.074
Both scores observed	738	889	1,060	2,687	
% of students	78.8	79.2	85.8	81.6	

Notes. Randomization sample; standard deviations in parentheses. The p-values test equality across the three arms, as described in the text.

arms ($p = 0.029$) but do not reject equal endline response rates ($p = 0.425$). Appendix Table A.3 compares covariate balance across all three arms in the randomized sample and among endline respondents. Baseline-score differences remain small, while gender and grade imbalances persist. These comparisons guide the inclusion of control variables in our intent-to-treat Equation (1) in Section 4.1.

3.3 Platform records

The platform panel records student-level activity by cycle for the classes with available platform data. A cycle's practice consists of base assignments, the diagnostic assignments of step (2), and the personalized assignments of step (4) that follow them. There are about 2.4 base and 1.9 personalized assignments per cycle, with 16 questions per assignment on average. Practice, $e_{it} \in [0, 1]$, is the share of the questions on the cycle's base and personalized assignments that student i attempted in cycle t. It measures how much of the assigned work a student attempts. The platform does not record the time or mental effort a student devotes to each question. Tournament attendance a_{it} indicates whether student i was present for the in-class tournament in cycle t; scores and recognition apply only to attending students.

The resulting student-cycle panel covers 2,317 students in 63 classes and 12 schools, with 7,951 student-cycles and 3.4 tournaments per student on average; the structural estimation uses the 7,372 student-cycles for which a baseline score is recorded. Two grade-6 team classes have no platform records, so the panel covers 63 of the 65 platform classes. The platform analyses use this panel or the relevant subset with observed baseline and endline scores. The main analysis also includes control students and students without platform records.

Table 2 describes the platform arms. Practice and tournament attendance barely differ between them. In both arms, students attempt about 68% of the practice questions assigned in a cycle and attend about two thirds of the tournaments. Raw tournament scores are somewhat higher in team classes, and by the construction of team recognition an attending student in a team class is more likely to be recognized.

Table 2. Platform activity by arm

	Individual	Team	Both arms
Students with platform records	1,121	1,196	2,317
Classes	30	33	63
Student-cycles	3,796	4,155	7,951
Practice, share of questions attempted	0.687	0.674	0.681
Tournament attendance	0.659	0.654	0.657
Tournament score, attending students	36.53	38.52	37.57
Recognized, attending students	0.268	0.420	0.347
Recognized, all students	0.177	0.275	
Students attending a tournament	26.9	25.4	
Teams per tournament		9.4	
Attending members per team		2.72	

Notes. Means over student-cycles of the platform arms; control classes held no tournaments.

4 Reduced-Form Evidence

This section reports on the effect of the experiment on end-of-year achievement and the behavior behind it. These facts will be the basis for the structural model that follows.

4.1 Treatment effects on achievement

We estimate the intent-to-treat (ITT) effects of the two platform arms on the endline score, relative to control,

$$S_i^{\text{end}} = \delta_0 + \lambda\, S_i^{\text{base}} + b_I\, \text{Individual}_i + b_T\, \text{Team}_i + \Gamma' X_i + \phi_{s(i)} + \varepsilon_i, \tag{1}$$

where i indexes students and $s(i)$ is the school of student i, δ_0 is the intercept, S_i^{end} is the endline score, S_i^{base} is the baseline score, Individual_i and Team_i are treatment indicators for the two platform conditions, and $\phi_{s(i)}$ is a school fixed effect. X_i contains explanatory variables including an indicator for a missing baseline, a female indicator, and grade indicators. The coefficients b_T and b_I are the ITT effects of being in a team-design and individual-design platform class relative to a control class.

Interpreting the arm contrasts as effects of assignment also requires that treatment in one class does not affect outcomes in another. Students remain with the same classmates during instruction, limiting some opportunities for contact across arms, but shared breaks and teachers provide possible channels for spillovers.[8] Teachers may teach classes in more than one arm, and time saved by the platform could affect their instruction elsewhere. To reduce variation in assignment preparation and content, we encouraged teachers to use the platform's question bank for paper assignments in control classes. These procedures limit some differences across arms but do not rule out spillovers.

Table 3 reports the estimation results for the 2,921 students with an endline score in 91 classes.[9] Standard errors are clustered by class, the unit of randomization, and to assess the sensitivity of inference to the finite number of classes, the table also reports p-values from the wild cluster bootstrap of Cameron et al. (2008) and from randomization inference that permutes arm labels across classes within school, the experiment's own assignment mechanism, over 2,000 draws (Young, 2019).

Table 3. Intent-to-treat effects on the endline score

		p-value			Effect
Contrast	Coef. (se)	CRVE	Wild	RI	in SD
Individual	+1.414 (1.277)	0.271	0.309	0.418	0.088
Team	+3.764** (1.325)	0.006	0.016	0.016	0.234
Team − Individual	+2.350 (1.313)	0.077	0.106	0.132	0.146
Students / classes	2,921 / 91				

Notes. The third row presents the difference $b_T - b_I$ from the same regression (1). Standard errors clustered by class. The p-values are cluster-robust (CRVE), from the wild cluster bootstrap (Wild) and from randomization inference (RI). Stars refer to RI, with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The team effect is about 3.76 points (0.23 SD) and is statistically significant under all three inference methods. The individual-arm effect, 1.41 points, is positive but imprecise, with p-values between 0.27 and 0.42. The team–individual difference, 2.35 points (0.15 SD),

[8] Students attend classes for all subjects with the same classmates in Indian schools.

[9] Endline assessments could not be done in one individual class, so the endline sample has 91 of the 92 classes.

is suggestive, with marginal statistical significance. These results are robust over a series of checks, which we present in Appendix Table A.4, with the individual-arm estimate between 0.9 and 1.5 points, the team-arm estimate ranging between 2.09 and 4.27 points, and the team–individual difference remaining positive. We therefore summarize our first main insight as follows.

Fact 1. *Students in team classes score higher at endline assessment than students in control classes. Students in individual classes score higher than the control group as well, but to a lesser extent than those in team classes, and the difference is imprecisely estimated.*

The small and imprecise gain in individual classes may reflect the platform's rudimentary personalization, under which students see their performance by topic after the diagnostic assignment, with links to videos on the topics where they did poorly, and must learn from the videos on their own. In team classes, teammates can help students fill in their knowledge gaps by explaining what they do not understand. The remaining evidence examines whether observed helping is consistent with this explanation.

4.2 Who helps whom

Within team classes, students are assigned to teams by drawing members from baseline-score quartiles. This procedure makes teammate assignment exogenous conditional on the formation protocol and balances baseline ranks across teams. It does not randomize subsequent practice, diagnostic reports, or helping. The regressions below, therefore, describe conditional relationships among these behaviors.

Recall that in each cycle, students go through two assignments, a base assignment and a personalized assignment, and the platform provides feedback on the current cycle's base assignment performance. In the individual-design, students observe only their own performances, while in the team-design, students observe the performances of all their team members before working on their personalized ones. Help between teammates is recorded on personalized assignments. Specifically, the last item of each personalized assignment asks the student to name the teammates who helped with it. This information can then be used to infer the help that each student receives and gives in each cycle. On average, a student gives help 0.72 times per cycle.

To examine how teammates help each other, we define a teammate as shown "behind" when their score on the base assignment of the cycle is below the median among team members who did the base assignment. We compute the indicator for each teammate as shown "behind," based on detailed information about team members' performance in each

cycle's base assignment. We then estimate

$$O_{jt} = b_L\, F^L_{jt} + b_F\, F_{jt} + b_A A_{jt} + \gamma\,(S^{\text{base}}_j - \bar{S}^{\text{base}}_{-j}) + \phi_{c(j)t} + \varepsilon_{jt}, \tag{2}$$

where O_{jt} is the number of reported instances of help given by j during cycle t, F^L_{jt} is the number of j's teammates shown behind, F_{jt} is the number of teammates who worked the cycle's personalized assignment, A_{jt} is a dummy variable that indicates j did the personalized assignment, $S^{\text{base}}_j - \bar{S}^{\text{base}}_{-j}$ is j's baseline score minus the mean baseline of j's teammates, and $\phi_{c(j)t}$ is a class-by-cycle fixed effect. The coefficient of interest is b_L. We control for the number of teammates working since more teammates at work means more people to help whatever the platform reports about them. For the help a student receives, we estimate

$$H_{jt} = b^H_I\, I_{jt} + b^H_F\, F_{jt} + b^H_e\, e^P_{jt} + \gamma^H(S^{\text{base}}_j - \bar{S}^{\text{base}}_{-j}) + \phi_{c(j)t} + \varepsilon_{jt}, \tag{3}$$

where H_{jt} is the number of reported instances of help received by j during cycle t, I_{jt} equals one when j is shown behind, and e^P_{jt} is the share of the cycle's personalized questions that j attempted, held fixed because help is associated with the questions a student works. We restrict this latter analysis to student-cycles in which the student tried the base assignment so that there is platform feedback about the student, because the feedback matters for the help that the student receives.

Table 4. Help and the platform's reports

	(1) Help given	(2) Help received
Teammates shown behind, F^L_{jt}	0.099***	
	(0.026)	
Student shown behind, I_{jt}		0.069***
		(0.025)
Teammates working, F_{jt}	0.127***	0.026
	(0.027)	(0.031)
Own personalized assign. tried	Yes	No
Relative baseline	Yes	Yes
Own share of personalized questions attempted	No	Yes
Class-by-cycle fixed effects	Yes	Yes
Observations	3,891	3,262

Notes. Column (1) estimates equation (2); column (2) estimates equation (3) on the student-cycles with platform feedback. Standard errors clustered by class. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4 reports the estimates. Column (1) shows that, holding fixed the number of teammates working, each additional teammate shown behind is associated with an increase in help given by 0.099 (0.026) per cycle. This is consistent with help being directed toward

teammates with diagnosed knowledge gaps. Unobserved changes in need or engagement could also contribute to this relationship. The results on received help in column (2) are also consistent, where a student shown behind receives 0.069 (0.025) more help per cycle than other students with a given share of the personalized assignment attempted. Appendix Table B.1 shows that in closely related specifications with student fixed effects, both associations are close to zero, 0.002 (0.025) for help given and 0.013 (0.023) for help received. Both associations, therefore, come from differences across students. Students whose teammates are more often shown behind give more help, and students who are more often shown behind receive more help. We summarize the insights from this analysis below.

Fact 2. *Help within teams is associated with platform's diagnostic feedback. Holding fixed the number of teammates working, a student gives more help the more teammates the platform shows to be behind, and a student shown behind receives more help.*

Finally, we find no statistically detectable relationship between helping and the team's previous competitive position. In a regression of help given on indicators for the team's rank in the previous tournament, controlling for own practice and class-by-cycle fixed effects, the rank indicators are jointly insignificant, $F = 0.45$ ($p = 0.773$). Relative to teams ranked sixth or lower, the estimated helping difference is 0.044 (0.037) for teams previously in the top three and 0.009 (0.037) for teams just outside them. These estimates do not establish that recognition incentives are absent, but they provide little evidence of a response to this measure of competitive position.

4.3 Collaboration and tournament performance

Tournament scores for students are observed up to seven times a year, which allows us to relate collaboration within a team to tournament performance, keeping fixed everything about the student that does not change across cycles. We define collaboration as the number of reported incidents of help in which the student is involved during a cycle, receiving and giving added together, $C_{it} = H_{it} + O_{it}$, where H_{it} is the number of help i received and O_{it} the number of help i gave in cycle t. The idea is that receiving help can improve a student's performance, and giving help can also improve the helper's knowledge. We therefore estimate

$$Y_{it} = \phi_i + \phi_{c(i)t} + b_e^Y \, e_{it} + b_F^Y \, F_{it} + b_c^Y \, C_{it} + \varepsilon_{it}, \tag{4}$$

where Y_{it} is student i's score in tournament t, e_{it} is the share of the cycle's assignment questions attempted, F_{it} is the number of teammates working the cycle's assignments alongside the student, ϕ_i is a student fixed effect and $\phi_{c(i)t}$ is a class-by-tournament fixed effect, so the

comparison is within a student across cycles, relative to the classmates sitting in the same tournament. We estimate this relationship among students in the team arm because there is no collaboration in the individual arm.

Table 5. Tournament performance, practice and collaboration in the team arm

	(1)	(2)
Practice e_{it}	2.156	0.892
	(1.621)	(1.719)
Teammates working F_{it}		0.308
		(0.805)
Collaboration C_{it}		0.565***
		(0.200)
Student fixed effects	Yes	Yes
Class-by-tournament fixed effects	Yes	Yes
Observations	2,719	2,719

Notes. Both columns use students in the team arm. Column (1) includes practice alone, and column (2) adds teammates working and collaboration as in equation (4). Standard errors in parentheses are clustered by class. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table 5 reports the estimation results. In column (1), the coefficient on practice alone is positive but imprecise at 2.156 (1.621). Column (2) adds teammates working and collaboration. Each additional incident of collaboration is associated with 0.565 (0.200) more tournament-score points. Student fixed effects absorb permanent differences across students. They do not remove time-varying differences in engagement, difficulty, or learning needs that may affect both collaboration and scores. Appendix Table B.2 reports an interaction specification and a specification separating help received from help given. The interaction between practice and collaboration is not statistically significant, and the coefficients for receiving and giving help cannot be distinguished from one another ($p = 0.824$). Additional robustness checks are reported in Appendix Table B.3. We summarize these results below.

Fact 3. *Holding fixed a student's own practice and the number of teammates working alongside them, tournament scores are positively associated with the collaboration a student takes part in during the cycle.*

We next examine whether average tournament performance predicts endline achievement (i.e., math score), controlling for annual practice and student characteristics. We estimate

$$S_i^{\text{end}} = \delta_0 + b_Y \bar{Y}_i + b_E E_i + \lambda S_i^{\text{base}} + \Gamma' X_i + \phi_{s(i)} + \phi_{g(i)} + \varepsilon_i, \tag{5}$$

where $\bar{Y}_i$ is student i's average score across attended tournaments, and E_i is the share of all

assigned base and personalized questions attempted over the year. The student characteristics are the baseline score S_i^{base} and X_i, which contains indicators for the missing baseline score and being female. The terms $\phi_{s(i)}$ and $\phi_{g(i)}$ are school and grade fixed effects. The regression uses one observation per student across the two platform arms, with standard errors clustered by class. Each student has only one endline outcome, so the regression controls for observed student characteristics and does not include student fixed effects.

Table 6. Endline achievement, annual practice and tournament performance

	Both platform arms
Annual practice E_i	5.495^{*}
	(2.918)
Average tournament score $\bar{Y}_i$	0.260^{***}
	(0.040)
Baseline, missing indicator, female	Yes
School and grade fixed effects	Yes
Students	1,880

Notes. The outcome is the endline score. Standard errors in parentheses are clustered by class. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table 6 reports the estimates. A one-point higher average tournament score is associated with 0.260 (0.040) more endline points, conditional on annual practice, baseline achievement, gender, school, and grade. This analysis leads to the following statement.

Fact 4. *A student's average tournament score is positively associated with endline achievement, conditional on annual practice, baseline achievement, gender, school, and grade.*

Together, Facts 3 and 4 show that collaboration is positively associated with tournament scores within students, and average tournament performance predicts endline achievement across students. Tournament scores serve as a measure of learning during the year. So, there is a plausible path for collaboration to contribute to learning that is ultimately reflected in endline achievement. These associations support the proposed learning channel but do not establish a causal effect of collaboration on endline achievement.

4.4 Practice and tournament attendance

In this section, we examine whether the positive team effect relative to control is due to more practice or higher tournament attendance. To assess this channel, we compare the two platform arms using a student-cycle level regression,

$$W_{it} = \delta_0 + b\,\text{Team}_i + \lambda S_i^{\text{base}} + \Gamma' X_i + \phi_{s(i)} + \varepsilon_{it}, \tag{6}$$

where W_{it} is practice e_{it} or tournament attendance a_{it}, equal to one if student i attends tournament t; Team_i equals one for students in team classes, and individual classes are the comparison group, so b is the difference between the two arms. The remaining terms are as in equation (1). Control students do not appear because practice and attendance are recorded only on the platform.

Table 7. Practice and attendance in team versus individual classes

	(1) Practice	(2) Attendance
Team class, b	−0.018	0.014
	(0.015)	(0.022)
Baseline, baseline-missing, female	Yes	Yes
School and grade fixed effects	Yes	Yes
Observations	7,951	7,951
Classes	63	63

Notes. Standard errors clustered by class. $^{***}\, p < 0.01$, $^{**}\, p < 0.05$, $^{*}\, p < 0.10$.

Table 7 reports small and statistically insignificant differences in average practice and attendance. These results fail to provide any evidence that the team design increases either outcome on average.

Overall, the reduced-form results identify a positive team effect (relative to the control) on endline achievement, while the team-versus-individual contrast is imprecisely estimated (Fact 1). Table 7 shows that the two platform arms have similar average practice and attendance. Within teams, help is associated with the platform's diagnostic feedback and with the number of teammates working (Fact 2). Collaboration is positively associated with tournament scores within students, and tournament scores are positively related to endline achievement.

Note that the team-versus-individual contrast combines the effects of team formation, sharing the platform's diagnostic reports among teammates, and team recognition. Team arm does not vary whether the reports are shared among teammates while holding teams and the remaining platform functions fixed. To disentangle these effects, we propose and estimate a structural model and run counterfactual analyses in the following.

5 Structural Model

5.1 Overview and timing

The model describes a school year consisting of several similarly structured cycles. Within a cycle, students choose what share of their assigned questions to attempt; the platform turns completed assignments into information about knowledge gaps; teammates decide whether to help one another; and students decide whether to attend the tournament. Students differ in how productive their practice is and in the cost of their study time. Each decision anticipates the stages that follow it, so the cycle is solved backward and is presented in the same backward order, in which the last decision, attendance at the tournament, comes first, then helping, then practice.

Three reduced-form findings guide the model's main assumptions. First, helping is associated with the platform's diagnostic feedback (Fact 2), motivating the assumption that students value helping teammates whose feedback indicates learning gaps. Second, collaboration is positively associated with tournament performance conditional on practice (Fact 3), so the model allows collaboration to produce knowledge alongside a student's own practice. Third, average tournament performance is positively associated with endline achievement (Fact 4), motivating the assumption that assessments reflect the knowledge students acquire during the year.

5.2 Knowledge, recognition and attendance at the tournament

We begin with the last decision. After practice and helping are done, student i decides whether to attend tournament t. Attending is valuable for the chance of being recognized, so we first specify what determines the tournament score and how recognition is awarded, and then the attendance decision.

Knowledge and the tournament score

The tournament score an attending student obtains is

$$Y_{it} = \alpha_0 + \alpha_z z_i + k_{it} + \sigma_\alpha \omega_i + \sigma_Y \varepsilon_{it}^Y, \tag{7}$$

where z_i is the standardized baseline score, k_{it} is the knowledge the tournament reads, α_0 is an intercept, ω_i is a permanent student effect in test performance randomly drawn from a standard normal distribution, and ε_{it}^Y is a standard normal transitory score shock. Here, σ_α is the standard deviation of the permanent component of tournament-score variation

and σ_Y that of the transitory component. The knowledge k_{it} comes from practice and from collaborating with a teammate. Let $e_{it} \in [0, 1]$ be the share of the cycle's assignment questions that student i attempts (i.e., practice), and C_{it} the number of collaborations i is involved in during cycle t (i.e., help received plus help given); both practice and collaboration are related to decisions taken earlier in the cycle and are specified below. The knowledge the tournament of cycle t reads is

$$k_{it} = \theta_i \, e_{it} + \zeta \, C_{it}, \tag{8}$$

where θ_i is the productivity of student i's practice, the knowledge produced by attempting every assigned question of a cycle, and ζ is the knowledge one collaboration produces. Productivity is a permanent type,

$$\log \theta_i = \theta_0 + \sigma_\theta u_i^\theta, \tag{9}$$

with u_i^θ standard normal, θ_0 the mean of log productivity, and σ_θ the permanent spread across students. A tournament tests the material of its own cycle, so it reads the practice and the collaboration of that cycle.[10]

Tournament Recognition

Recognition in a tournament follows a known recognition rule, which differs between the two platform arms. Under the individual recognition rule, used in individual classes, the five highest scores among attending students are recognized. Under the team recognition rule, used in team classes, the top three teams with the highest sum of attending members' scores are recognized, and only attending members of those teams are recognized. Appendix Table B.8 shows that the recorded outcomes follow the two rules and how ties raise the number of students recognized under the individual rule.

The team recognition rule has two consequences that the model uses. A member who attends adds their score to the team's standing, so attendance is itself a contribution to winning; and help that raises a teammate's score raises the team's total score by the full amount. The team recognition rule also makes a student's chance of recognition depend less on the student's own score than under the individual recognition rule, since a team can place in the top three while a member places mid-class, and the data show this flatter profile. Among attending students, recognition rises from 0.24 in the bottom quartile of

[10] Column (2) of Appendix Table B.3 shows that the associations of prior cycles' practice and collaboration with the current tournament score are statistically insignificant.

baseline score to 0.31 in the top under the individual recognition rule but lies between 0.39 and 0.45 in every quartile under the team recognition rule.

Because the recognition rule is known and the model specifies the distribution of scores through (7), the probability that student i is recognized in cycle t, denoted by EP, is a well-defined function of the knowledge of everyone in the class,

$$\widehat{\mathrm{EP}}_{it}^{\,\mathrm{ind}} = \Pr(\#\{\text{attending classmates scoring above } i\} < 5 \mid a_{it} = 1)\,, \tag{10}$$

$$\widehat{\mathrm{EP}}_{it}^{\,\mathrm{team}} = \Pr(\#\{\text{other teams with a total score above that of } i\text{'s team}\} < 3 \mid a_{it} = 1)\,, \tag{11}$$

Both probabilities condition on i attending, so that i's own score is in the team's total score. We write $\widehat{\mathrm{EP}}_{it}$ for whichever probability applies to i's class and call it the recognition probability; the notation stands for the expected prize, since recognition is the tournament's prize, and the hat marks a probability computed from the model. Let $u(i)$ denote student i's team and $T_{ut} = \sum_{j \in u} a_{jt} Y_{jt}$ denote team u's total score in cycle t, where a_{jt} indicates tournament attendance. The total score of student i's team is therefore $T_{u(i),t}$.

Students form the above probabilities before the tournament. They know every classmate's expected score, $\alpha_0 + \alpha_z z_j + k_{jt}$, with the collaboration j takes part in at its expected level. Collaboration enters at its expected level because students do not observe the help given in other teams, and we maintain this for their own team as well. They do not know who exactly will attend, or the test-performance terms $\sigma_\alpha \omega_j + \sigma_Y \varepsilon_{jt}^Y$, their own included, which are integrated out, but they know the distribution of these variables.

Tournament scores in our data are recorded as the share of questions answered correctly, in points out of 100, so in a class-cycle whose tournament has Q questions the recorded scores move in steps of $g = 100/Q$. As a result, there is a non-negligible chance of ties happening in tournaments, as shown in Table B.8. Because the scores in the model are continuous while the recorded ones move in steps, we need to model the possibility of ties and derive the recognition probabilities from these primitives. We present these derivations in Appendix D.1. The decisions below use two derivatives of $\widehat{\mathrm{EP}}_{it}$, one with respect to i's own knowledge k_{it} and, under the team recognition rule, one with respect to a teammate's knowledge k_{jt}, which enters the team's total score only when j attends.

The tournament attendance decision

Attending a tournament has a cost in time. The cost of student i's time in cycle t is

$$\log c_{it} = \pi_b z_i + \sigma_\nu \nu_{it}, \tag{12}$$

where π_b measures how much of the cost of study time is predicted by the baseline score and ν_{it} is a standard normal shock with spread σ_ν, drawn independently each cycle. We use the same cost of time in the tournament attendance decision and the practice decision (specified below), so that a higher cost in a given cycle discourages both activities. Appendix Table B.6 shows a positive within-student association between practice and attendance, which motivates this shared cost. The cycle-specific shock allows both decisions to vary over time. Appendix Table B.7 shows that in the data, within-student variation accounts for 36% of the variance of log practice in cycles with $0 < e_{it} < 1$ and 54% of the variance of attendance. The permanent differences in practice across students come from productivity θ_i and from baseline score through π_b.

We define the utility of attending a tournament as

$$u_{it}^A = \mu_a - \rho_c \log c_{it} + B_v \left(\widehat{\mathrm{EP}}_{it} - \overline{\mathrm{EP}}\right) + \varepsilon_{it}^1, \tag{13}$$

and normalize the systematic utility of not attending to zero, so that $u_{it}^0 = \varepsilon_{it}^0$, where ε_{it}^1 and ε_{it}^0 are independent type-I extreme-value shocks. Here, μ_a is the attendance intercept, ρ_c is the loading on the cost of time, and B_v is the value of recognition. The recognition probability $\widehat{\mathrm{EP}}_{it}$ is conditional on student i attending, and $\overline{\mathrm{EP}}$ is its mean over the student-cycles of the two platform arms. Thus, μ_a is the systematic utility of attending for a student with the average recognition probability and $\log c_{it} = 0$. Centering gives μ_a this meaning, since without it μ_a would be the utility of attending for a student with no chance of recognition, and score shocks give every attending student a positive chance. It also separates the roles of the two parameters, so that μ_a governs the attendance rate and B_v governs how attendance varies with the recognition probability around that rate. Because $\overline{\mathrm{EP}}$ is the same number for every student and cycle, centering changes only the intercept and leaves the attendance probabilities unchanged.

For the attendance probability, we use the systematic component of utility,

$$\bar{u}_{it}^A \equiv u_{it}^A - \varepsilon_{it}^1 = \mu_a - \rho_c \log c_{it} + B_v\big(\widehat{\mathrm{EP}}_{it} - \overline{\mathrm{EP}}\big). \tag{14}$$

The attendance shocks are integrated out when calculating this probability. The attendance decision is forward-looking within the cycle. Before tournament scores and recognition are realized, the student weighs the cost of attending against the anticipated benefit of recognition, using $\widehat{\mathrm{EP}}_{it}$ to account for uncertainty about scores and other students' attendance. Student i attends the tournament if $u_{it}^A \geq u_{it}^0$. Writing $a_{it} = 1$ for attendance, the decision rule implies

$$\Pr(a_{it} = 1) = \Lambda(\bar{u}_{it}^A), \tag{15}$$

where Λ is the logistic function. The expected value of the attendance stage before the shocks are drawn is

$$V_{it}^{A} = \log\bigl(1 + e^{\bar{u}_{it}^{A}}\bigr), \qquad \frac{\partial V_{it}^{A}}{\partial \bar{u}_{it}^{A}} = \Lambda(\bar{u}_{it}^{A}), \tag{16}$$

the expected maximum over attending and not, up to an additive constant independent of the decisions. V_{it}^{A} is the continuation value that every earlier decision in the cycle carries.

5.3 Information

Working backward, the decision before attendance is helping within teams, and it responds to what teammates know about one another. The platform provides each student with feedback on the assignment, and in team classes, teammates can see one another's feedback. The data records the assignment score associated with that feedback, and we summarize the information a teammate's report conveys into a single indicator. Student k is shown behind, $I_{kt} = 1$, when k did the cycle's assignment and k's base score for the cycle is below the median score among participating team members, including k. The researchers constructed the below-median indicator from diagnostic scores, and the platform did not randomize it as a separate message. The indicator is constructed from recorded scores for each student-cycle and is entered into the model as an observed state.

5.4 Helping

Helping occurs after the assignment feedback and before the tournament attendance decision. We model the probability that student j helps teammate k using a logistic function, scaled by the share of the assignment that k attempts,

$$\Pr(h_{jkt} = 1) \;=\; e_{kt}\,\Lambda\!\left(\tau\, I_{kt} \;+\; \zeta\left[\Lambda(\bar{u}_{jt}^{A})\, B_v\left(\frac{\partial \widehat{\mathrm{EP}}_{jt}}{\partial k_{jt}} + \frac{\partial \widehat{\mathrm{EP}}_{jt}}{\partial k_{kt}}\right) \;+\; \beta^{e}\right] \;-\; \kappa_h\right), \tag{17}$$

where h_{jkt} equals one when teammate j helps student k in cycle t, and Λ is the logistic function. The factor e_{kt} outside the logistic function scales the probability by the share of the assignment that recipient k attempts. It makes help impossible when k does no work and allows more practice to create more opportunities for help. In the data, help received is exactly zero in every team-arm student-cycle with no work done, and column (7) of Appendix Table B.6 shows that it rises with the share attempted. Inside the logistic function, the first term, τI_{kt}, captures the response to the platform's diagnostic feedback. When the platform shows k behind, $I_{kt} = 1$, this term adds τ to the helping index. A positive τ therefore increases the probability of helping a teammate shown behind, holding

the other terms fixed. The next term, $\zeta\Lambda(\bar{u}^A_{jt})B_v\big(\partial\widehat{\mathrm{EP}}_{jt}/\partial k_{jt} + \partial\widehat{\mathrm{EP}}_{jt}/\partial k_{kt}\big)$, captures the recognition benefit of helping. One collaboration adds ζ to each student's knowledge by (8), and both gains enter the same team total score $T_{u(j),t}$. The two derivatives convert them into a change in j's recognition probability. Conditional on j attending, j's own gain enters the total score for certain, and k's gain enters it only when k attends, which the derivative with respect to k_{kt} takes into account. Multiplying by B_v values that change, and multiplying by $\Lambda(\bar{u}^A_{jt})$ accounts for the fact that j receives recognition only if attending, as in (16).

The term $\zeta\beta^e$ captures the helper's own learning benefit. Student j gains ζ units of knowledge from explaining the material and values each unit at β^e. This benefit does not require tournament attendance. Finally, $-\kappa_h$ captures the cost of helping and lowers the helping probability. Because the learning and recognition benefits enter separately, κ_h represents the gross cost of helping. In individual classes, there are no teammates and no help. Counting over teammates,

$$H_{it} = \sum_{j \neq i} h_{jit}, \qquad O_{it} = \sum_{k \neq i} h_{ikt}, \qquad C_{it} = H_{it} + O_{it}, \tag{18}$$

where H_{it} is the number of teammates who helped i in cycle t, O_{it} is the number of teammates i helped, and C_{it} is the collaboration count introduced in (8).

5.5 Practice

At the start of each cycle, student i chooses what share of the assigned questions to attempt, $e_{it} \in [0, 1]$, a single decision over the cycle's assignments. We refer to this choice as practice. It measures the extent of the work attempted, since the time and mental effort devoted to each question are not observed. The student chooses practice anticipating the help it will attract and, through the continuation value (16), the tournament outcome, so that

$$e_{it} = \arg\max_{e \in [0,1]} \; \beta^e\big(\theta_i\, e + \zeta\, \mathbb{E}[C_{it} \mid e]\big) \; + \; V^A_{it}(e) \; - \; c_{it}\Big(\frac{N_{it}}{\bar{N}}\Big)^{\psi}\Big(\kappa_1 e + \tfrac{e^{\xi}}{\xi}\Big). \tag{19}$$

The first term on the right-hand side values the knowledge produced by own practice, $\theta_i e$, and expected collaboration, $\zeta\mathbb{E}[C_{it} \mid e]$, where expected collaboration follows (17). The parameter β^e is the value of a unit of knowledge relative to the cost of time; knowledge enters tournament and end-of-year scores at rates normalized to one. The second term, $V^A_{it}(e)$, is the expected value of the tournament attendance stage. Practice affects this value through knowledge, and hence the probability of recognition. The final term is the cost of practice. Here, N_{it} is the number of assigned questions and $\bar{N}$ its geometric mean over

student-cycles. The factor $(N_{it}/\bar{N})^{\psi}$ makes attempting a given share more costly in longer assignments when $\psi > 0$. This is motivated by data patterns that show within a student, the share of attempted questions falls by 0.071 per log point of assignment length, and full completion of an assignment is more common in shorter assignments (Appendix B.2). The parameter κ_1 is the slope of the linear component of the cost, and $\xi > 1$ governs the curvature of the convex component of the cost.

In choosing practice, the student takes classmates' practice and attendance probabilities as given and holds the helping indices fixed, including their own attendance probability within those indices. Thus, increasing e attracts more help through the factor e in (17), while help given remains unchanged. Each collaboration it attracts also adds ζ to the knowledge of the teammate who helps, which under the team recognition rule raises the team's total score. The marginal value of practice is

$$\begin{aligned} mv_{it} = \Big(\theta_i + \zeta\,\frac{\partial \mathbb{E}[C_{it} \mid e]}{\partial e}\Big)\Big(\beta^{e} + \Lambda(\bar{u}^{A}_{it})\,B_v\,\frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial k_{it}}\Big) \\ + \Lambda(\bar{u}^{A}_{it})\,B_v\,\zeta \sum_{j \neq i} \Lambda(\text{index of } j \text{ helping } i)\,\frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial k_{jt}}, \end{aligned} \tag{20}$$

where, in the first line, the first factor is the additional knowledge from own practice and the help it attracts, and the second is its value through learning and recognition. The derivative $\partial \mathbb{E}[C_{it} \mid e]/\partial e = \sum_{j \neq i} \Lambda(\text{index of } j \text{ helping } i)$ is the help received per unit of practice. The second line is the recognition value of the knowledge that the teammates who help i gain, ζ per collaboration; it is zero in individual classes, where there is no helping.

Differentiating the cost term gives the marginal cost of practice,

$$mc_{it}(e) = c_{it}\Big(\frac{N_{it}}{\bar{N}}\Big)^{\psi}\big(\kappa_1 + e^{\xi - 1}\big). \tag{21}$$

The linear component contributes $c_{it}(N_{it}/\bar{N})^{\psi}\kappa_1$ even when nothing is attempted, and the convex component adds $c_{it}(N_{it}/\bar{N})^{\psi}e^{\xi-1}$, which rises with the share of questions attempted. At an interior optimum of (19), marginal value equals marginal cost, $mv_{it} = mc_{it}(e_{it})$. Solving this equality for the share of questions attempted and incorporating the bounds $e_{it} \in [0, 1]$ gives the decision rule

$$e_{it} = \min\left\{ \left[\max\left(\frac{mv_{it}}{c_{it}(N_{it}/\bar{N})^{\psi}} - \kappa_1,\ 0 \right) \right]^{1/(\xi-1)},\ 1 \right\}, \tag{22}$$

$$e_{it} = 0 \iff \frac{mv_{it}}{c_{it}(N_{it}/\bar{N})^{\psi}} \leq \kappa_1, \tag{23}$$

with marginal value evaluated at the chosen share. A student attempts nothing when marginal value does not exceed marginal cost at $e = 0$, and attempts every assigned question when marginal value is at least marginal cost at $e = 1$. These bounds accommodate the masses of observations at $e_{it} = 0$ and $e_{it} = 1$ that Appendix Table B.5 reports. Appendix D.2 presents a step-by-step derivation of the above conditions.

Productivity and the cost of time enter the decision mainly through θ_i/c_{it}, so practice alone cannot separately identify high productivity from low costs. Tournament and end-of-year scores help separate them because productivity determines how much knowledge practice produces, whereas the cost of time does not.

5.6 Skill production

At the end of the academic year, students take an end-of-year assessment. We specify the score on this assessment as

$$S_i^{\text{end}} = \delta_0 + \lambda S_i^{\text{base}} + \delta_C \,\text{Control}_i + \textstyle\sum_t \theta_i e_{it} + \zeta \sum_t C_{it} + \gamma_{s(i)} + \gamma_{g(i)} + \varepsilon_i, \tag{24}$$

where δ_0 is a common intercept, the indicator Control_i equals one for control students and zero for students in the individual and team arms, $\gamma_{s(i)}$ and $\gamma_{g(i)}$ are fixed effects for student i's school and grade, and ε_i is a normally distributed residual, i.i.d. across students and independent of school and grade, with mean zero and standard deviation σ_ε. Every student, in every arm, has a school fixed effect and a grade fixed effect. They capture differences in end-of-year scores across schools and grades that have nothing to do with the platform. We include them because the arms differ in grade composition, with 14 of the 22 grade-6 classes in the team arm (Section 3.2), and without them the model would attribute grade differences in end-of-year scores to the platform. We measure the fixed effects relative to the average control student, so they average zero among control students. The fixed effects are not among the parameters of the model, and we handle them as in the intent-to-treat regressions of Section 4.1. Each comparison of end-of-year scores that we use comes from a regression that includes school and grade indicators. The coefficients on these indicators absorb the fixed effects, and the remaining coefficients of the regression do not depend on them. Knowledge enters at a rate of one by the normalization above. The end-of-year assessment covers everything taught during the year, so it reads all practice and all collaboration, the same stock that the tournament reads cycle by cycle.

Control students have no platform records, so for them the practice and collaboration sums are zero and their end-of-year score is $\delta_0 + \delta_C + \lambda S_i^{\text{base}} + \gamma_{s(i)} + \gamma_{g(i)} + \varepsilon_i$. Control students still learn during the year, from classroom instruction and paper assignments that

the platform does not record, and δ_C captures this learning. The end-of-year difference between a platform arm and control, therefore, compares the knowledge built on the platform and in team classes through collaboration, with what control students learn away from it.

5.7 Equilibrium

Students interact in two ways. In both platform arms, a student's chance of recognition depends on the knowledge and attendance of everyone in the class, so practice and attendance in a class form a game played through the tournament. Recognition depends on knowledge, knowledge on practice, and practice on the chance of recognition, and the attendance pool over which the tournament is played is itself the attendance decision of (15).

In team classes, students also interact through helping, because $\mathbb{E}[C_{it} \mid e]$ in (19) depends on what teammates do. Helping is a set of pairwise decisions within a team. In individual classes there are no teammates and $\mathbb{E}[C_{it} \mid e] = 0$, so the tournament is the only source of interaction. In both arms, each student's best response takes classmates' practice and attendance probabilities as given, and the equilibrium of a class in a cycle is a fixed point of the decision rules above. We solve for the fixed point at every parameter vector by best-response iteration. Across cycles, decisions are assumed to be separable, in that the knowledge of every cycle reaches the end-of-year assessment and nothing a student does in one cycle changes the payoffs of the next. Columns (5) and (6) of Appendix Table B.6 show insignificant associations of prior recognition with subsequent practice and attendance, which inspire this assumption.

6 Identification

The model has twenty-two parameters. Three govern productivity and the knowledge produced by collaboration (θ_0, σ_θ, ζ), six govern the value and cost of practice (β^e, ξ, κ_1, ψ, π_b, σ_ν), three govern attendance and the value of recognition (μ_a, ρ_c, B_v), two govern helping (τ, κ_h), four govern the tournament score (α_0, α_z, σ_α, σ_Y), and four govern end-of-year achievement (δ_0, λ, δ_C, σ_ε). We identify them from twenty-two moments of the data. The assumptions and propositions below state the conditions under which these moments identify the parameters.

Identification proceeds in three steps that follow the design of the experiment. The first step uses the individual arm, where there is no helping, so that tournament scores, practice and attendance depend only on the technology and on the behavioral parameters. One moment of this step, the within-student slope of the tournament score on practice, uses both

platform arms and includes the number of collaborations as a covariate. The second step adds the end-of-year scores of control students, which pin down the end-of-year production function up to the level of what students learn away from the platform. The third step adds the end-of-year contrasts of the two platform arms with control and the helping data of the team arm, which identify that level, the knowledge produced by collaboration and the helping decision. Each step takes as given the parameters recovered in the steps before it. Table 8 lists, for each parameter, the moment that determines it and the step in which it is recovered. The rest of this section shows why each moment identifies the parameter it is paired with.

Four assumptions underlie the identification argument. They concern the experimental design, the distribution of the unobserved shocks, the recognition rule and the support of the observed variables.

Assumption 1 (Design)**.** *Experimental arms are assigned within school, students to teams within class, and assignment lengths N_{it} and tournament lengths Q_{ct} by the platform, independently of $(\theta_i, \{\nu_{it}, \varepsilon^Y_{it}, \varepsilon^h_{jkt}\}, \omega_i, \varepsilon_i)$ given z_i.*

Arms were randomly assigned to classes within schools, and students were randomly assigned to teams within their class, one from each baseline quartile, so the assumption holds by design for arms and team membership. Independence of assignment length from latent productivity and cost shocks is an additional modeling assumption. Two facts in Appendix B.2 support it. The base assignment is identical for every student in a class and cycle in 99% of class-by-cycle cells, so its length is set by the curriculum and cannot respond to any one student's productivity or cost shocks. The length of the personalized assignment does not vary with the baseline score and barely responds to how much of the base assignment the student attempted in the same cycle, so it does not follow the cost shock that drives practice in that cycle. Because diagnostic reports are observed after practice, their independence from the helping taste shocks conditional on the modeled state is also required to interpret τ structurally.

Assumption 2 (Shocks)**.** *$u^\theta_i, \nu_{it}, \varepsilon^Y_{it}, \omega_i, \varepsilon_i$ are normal with the stated scales and ε^h_{jkt} is standard logistic; all are mutually independent and independent of z_i, the arm and team membership.*

The normal distributions specify productivity and continuous shocks. Standard logistic helping shocks and the type-I extreme-value attendance shocks fix the utility scales of the discrete choices. Independence separates learning productivity, study costs, test performance,

and helping preferences. These restrictions are substantive because correlations among latent characteristics could provide alternative explanations for the observed moments.

Assumption 3 (Recognition rule)**.** *(i) Under the individual recognition rule the five highest attending scores are recognized, and so is every student tied with the fifth. (ii) Under the team recognition rule every attending member of the three teams with the highest total scores is recognized, ties included, where a team's total score is the sum of the tournament scores of its attending members. (iii) In a class-cycle whose tournament has Q_{ct} questions, recorded scores lie on a grid of spacing $100/Q_{ct}$. (iv) Students know the recognition rule of their class.*

Parts (i) to (iii) describe the platform's protocol, and the recorded outcomes are consistent with each of them (Appendix B.4). Under the individual recognition rule, ties at the fifth-highest score raise the number of students recognized to 7.2 per tournament on average. Under the team recognition rule, the ranking of teams by the total score of attending members has a Spearman correlation of 0.996 with the recorded team ranking. The spacing of recorded scores equals $100/Q_{ct}$ in 190 of the 195 class-cycles in which Q_{ct} is recorded. Part (iv) implies that students form their recognition probabilities under the rule that applies in their class.

Assumption 4 (Support)**.** *e_{it} has positive mass at 0, at 1 and on $(0,1)$; a positive mass of students have two or more cycles with distinct interior e_{it}, and a positive mass attend at least two cycles with distinct interior e_{it}; N_{it} varies within student across cycles; within team-cycles, I_{kt} varies across teammates with positive probability.*

These support conditions hold in the data. Appendix Table B.5 shows that of the 7,372 student-cycles, 8.3% have $e_{it} = 0$, 32.8% have $e_{it} = 1$ and the remaining 58.9% are interior. Of the 2,144 students, 1,423 have two or more cycles with distinct interior practice, and 913 attend at least two tournaments in such cycles. Assignment length differs across cycles for 98% of the students observed in more than one cycle, and variation within students accounts for 54% of the variance of log assignment length. In team classes, whether a student is shown behind differs across teammates in 86% of the 1,001 team-cycles. Under these assumptions, the propositions below state what identifies the parameters of each step. Some moments determine their parameter in closed form. For the others, the proof writes each moment as a known function of the parameters of its step and shows how the paired parameter moves it, and the moments of the step determine its parameters, in a neighborhood of their true values, when the Jacobian of these moments with respect to these parameters has full rank. Each proposition is followed by its intuition, and all proofs are in Appendix F.

Proposition 1 (Tournament scores, practice and attendance)**.** *Given the recognition probabilities and the recognition value of knowledge at their equilibrium values, the fifteen moments in the first two blocks of Table 8 identify the productivity distribution and the tournament score,* θ_0*,* σ_θ*,* α_0*,* α_z*,* σ_α *and* σ_Y*, and the nine parameters of practice and attendance,* β^e*,* ξ*,* κ_1*,* ψ*,* π_b*,* σ_ν*,* μ_a*,* ρ_c *and* B_v*.*

Consider first the tournament scores, taking the practice and attendance decisions as given. In individual classes, there is no collaboration, so the score of an attending student is $Y_{it} = \alpha_0 + \alpha_z z_i + \theta_i e_{it} + \sigma_\alpha \omega_i + \sigma_Y \varepsilon_{it}^Y$, where $\log \theta_i = \theta_0 + \sigma_\theta u_i^\theta$, and with practice e_{it} observed. Within a student, $\alpha_0 + \alpha_z z_i$ and ω_i do not change from cycle to cycle, so the within-student slope of the score on practice is an average of θ_i across students and rises with θ_0. We compute this slope on both platform arms, controlling for the number of collaborations, so that it uses all the variation in practice within students. Averaging the score equation gives $\mathbb{E}[Y_{it}] = \alpha_0 + \alpha_z \mathbb{E}[z_i] + \mathbb{E}[\theta_i e_{it}]$, and the slope of the score on z_i is α_z plus the slope of knowledge $\theta_i e_{it}$ on z_i, so the mean score and its slope on the baseline score give α_0 and α_z. A student's mean score over the T_i tournaments attended contains the permanent shock in full but the transitory shock with variance σ_Y^2/T_i, so $\mathrm{Var}(Y_{it}) = \mathrm{Var}(\alpha_z z_i + \theta_i e_{it}) + \sigma_\alpha^2 + \sigma_Y^2$ and $\mathrm{Var}(\bar{Y}_i) = \mathrm{Var}(\alpha_z z_i + \theta_i \bar{e}_i) + \sigma_\alpha^2 + \frac{\sigma_Y^2}{\bar{T}}$, where $\bar{e}_i$ is the student's mean practice over those tournaments and $\bar{T}$ is the mean number of tournaments attended per student. The variance of scores and its between-student share therefore separate σ_α from σ_Y.

Consider next practice, taking the tournament score parameters and the recognition value of knowledge as given. In individual classes the marginal value of practice is $mv_{it} = \theta_i(\beta^e + r_{it})$, where $r_{it} = \Lambda(\bar{u}_{it}^A) B_v \, \partial \widehat{\mathrm{EP}}_{it} / \partial k_{it}$ is the value of a unit of knowledge through recognition, and the cost of a unit of practice is $c_{it}(N_{it}/\bar{N})^\psi$ with $\log c_{it} = \pi_b z_i + \sigma_\nu \nu_{it}$. The decision rule (22) depends on their ratio $x_{it} = mv_{it}/\{c_{it}(N_{it}/\bar{N})^\psi\}$,

$$e_{it} = \begin{cases} 0 & \text{if } x_{it} \leq \kappa_1, \\ (x_{it} - \kappa_1)^{1/(\xi-1)} & \text{if } \kappa_1 < x_{it} < \kappa_1 + 1, \\ 1 & \text{if } x_{it} \geq \kappa_1 + 1, \end{cases}$$

and $\log x_{it} = \theta_0 + \sigma_\theta u_i^\theta + \log(\beta^e + r_{it}) - \pi_b z_i - \psi \log \frac{N_{it}}{N} - \sigma_\nu \nu_{it}$. Raising κ_1 raises the share with no work, and raising β^e raises x_{it} in every cycle and with it the share with full completion, so the two corner shares determine κ_1 and β^e. Between the corners, $\log e_{it} = \log(x_{it} - \kappa_1)/(\xi - 1)$ with $x_{it} - \kappa_1$ between zero and one, so interior log practice is negative and its mean moves toward zero as ξ rises, which determines ξ. The baseline score and assignment length enter $\log x_{it}$ with coefficients $-\pi_b$ and $-\psi$, so the slope of log practice on the baseline score determines π_b and the within-student slope of practice on log assignment length determines

ψ. The two dispersions are identified from interior practice. On an interior cycle, practice reveals the index, $\log x_{it} = \log(\kappa_1 + e_{it}^{\xi-1})$, and once its known parts are removed what remains is the composite shock $\sigma_\theta u_i^\theta - \sigma_\nu \nu_{it}$. Over all cycles this shock has variance $\sigma_\theta^2 + \sigma_\nu^2$, and its values in two cycles of the same student have correlation $\sigma_\theta^2/(\sigma_\theta^2 + \sigma_\nu^2)$, because the cycles share u_i^θ. A cycle is interior when the shock lies in a known interval, so interior practice reveals the distribution of the shock truncated to that interval, which depends on these two parameters. The spread of interior practice therefore determines the total dispersion, and the part of that spread which is common to a student's cycles determines how much of it is permanent, which gives σ_θ and σ_ν. The between- and within-student variances of interior log practice are the two summaries of this distribution that the estimator matches.

Consider finally attendance, taking the recognition probabilities as given. A student attends with probability

$$\Pr(a_{it} = 1) = \Lambda\big(\mu_a - \rho_c \log c_{it} + B_v(\widehat{\mathrm{EP}}_{it} - \overline{\mathrm{EP}})\big) = \Lambda\big(\mu_a - \rho_c \pi_b z_i - \rho_c \sigma_\nu \nu_{it} + B_v(\widehat{\mathrm{EP}}_{it} - \overline{\mathrm{EP}})\big),$$

where Λ is the logistic function. The probability rises with μ_a, so the attendance rate determines μ_a. The cost shock ν_{it} is the only transitory shock that enters both practice and attendance, and on an interior cycle practice reveals it, since $\log x_{it} = \log(\kappa_1 + e_{it}^{\xi-1})$ differs from the rest of the index by $\sigma_\nu \nu_{it}$. Within a student, the log-odds of attendance then rise with the revealed $\log x_{it}$ at the rate ρ_c, so the within-student slope of attendance on interior log practice determines ρ_c. The baseline score enters the attendance index in two ways only, through the cost of time, with coefficient $-\rho_c \pi_b$, and through the recognition probability $\widehat{\mathrm{EP}}_{it}$, which rises with the baseline score because the expected score $\alpha_0 + \alpha_z z_i + \theta_i e_{it}$ does. With ρ_c and π_b determined, the slope of attendance on the baseline score within class and cycle determines B_v. This is the exclusion restriction behind B_v. The baseline score shifts attendance only through the cost of time and the chance of recognition.

The three arguments depend on one another. The argument for scores takes practice as given, the argument for practice takes the value of recognition as given, which depends on the score parameters and on attendance, and the argument for attendance takes the cost parameters of practice and the recognition probabilities as given. The fifteen parameters are therefore identified together.

Proposition 2 (End-of-year achievement)**.** *The mean end-of-year score of control students and, net of school and grade, its slope on the baseline score and the standard deviation of its residual identify λ, σ_ε and the sum $\delta_0 + \delta_C$.*

Control students have no platform records, so their end-of-year score is $S_i^{\mathrm{end}} = \delta_0 + \delta_C + \lambda S_i^{\mathrm{base}} + \gamma_{s(i)} + \gamma_{g(i)} + \varepsilon_i$. In a regression of the score on the baseline score and on school and

grade indicators, the indicators absorb the two fixed effects, the coefficient on the baseline score is λ, because ε_i is independent of the baseline score, school and grade, and the residual has standard deviation σ_ε. The school and grade effects average to zero among control students, so the mean gives the sum $\delta_0 + \delta_C$. Separating δ_C, what control students learn away from the platform, from the intercept requires the platform arms, and this is done in the third step.

Proposition 3 (Collaboration, helping and the control-arm level)**.** *Given Propositions 1 and 2 and the practice of team students at its equilibrium value, the individual–control and team–control contrasts, the level of help given and the slope of help given on the number of teammates shown behind identify δ_C, ζ, τ and κ_h, and δ_0 then follows from Proposition 2.*

An individual-arm student's end-of-year score is $\delta_0 + \lambda S_i^{\text{base}} + \sum_t \theta_i e_{it} + \gamma_{s(i)} + \gamma_{g(i)} + \varepsilon_i$ and a team student's is $S_i^{\text{end}} = \delta_0 + \lambda S_i^{\text{base}} + \sum_t \theta_i e_{it} + \zeta \sum_t C_{it} + \gamma_{s(i)} + \gamma_{g(i)} + \varepsilon_i$, where the number of collaborations $\sum_t C_{it}$ is observed. Each platform arm carries the knowledge built on the platform in place of δ_C, and the team arm carries in addition ζ times the number of collaborations. The two end-of-year contrasts, the differences between each platform arm and control in mean end-of-year scores net of baseline, school and grade, are the moments that link the model to the experiment. Netting out the knowledge from own practice, which Proposition 1 determines from productivity and observed practice, the two contrasts are two linear equations in δ_C and ζ, which determine both. Turning to helping, the probability that j helps teammate k is $\Pr(h_{jkt} = 1) = e_{kt}\,\Lambda\big(\tau I_{kt} + \chi_{jkt} - \kappa_h\big)$, where $I_{kt} = 1$ when k is shown behind and, from (17), $\chi_{jkt} = \zeta\big[\Lambda(\bar{u}^A_{jt})B_v\left(\partial\widehat{\text{EP}}_{jt}/\partial k_{jt} + \partial\widehat{\text{EP}}_{jt}/\partial k_{kt}\right) + \beta^e\big]$ is j's benefit from helping k. The first term in the brackets is the recognition benefit, the rise in j's chance of recognition from the knowledge the collaboration adds to the team's total score, and the second is the value of the knowledge j gains by explaining the material. Both depend on ζ and on the parameters of Proposition 1. A higher κ_h lowers the probability of helping every teammate, and a higher τ raises the probability of helping a teammate shown behind, $\Lambda(\tau + \chi_{jkt} - \kappa_h)$, relative to one not shown behind, $\Lambda(\chi_{jkt} - \kappa_h)$. When χ_{jkt} and the recipient's practice are the same for every pair, the slope of help given on the number of teammates shown behind, in a regression that also includes the number shown ahead, measures the first probability times that practice, and the level of help given, which counts help to both kinds of teammates, then gives the second. The difference in their log-odds is τ, and κ_h follows from the second. When pairs differ, the level still falls with κ_h and rises with τ, and the slope is a known function of both. The four parameters are identified jointly, because collaboration enters knowledge, knowledge enters the recognition benefit of helping, and in team classes practice responds to the help it attracts.

The three propositions take the equilibrium objects as given, the recognition probabilities, the recognition value of knowledge and the practice of team students, because scores, practice, attendance and helping are decided in one equilibrium. Whether the twenty-two moments determine the twenty-two parameters jointly is a question about the full map from parameters to moments, in which the equilibrium responds to every parameter. We compute the equilibrium of each class and cycle by best-response iteration from a fixed starting point, which selects one equilibrium at each parameter vector, and we assume that the moments are continuously differentiable functions of the parameters at the selected equilibrium. The parameters are then locally identified, in the sense that no other parameter vector near the true one produces the same moments, when the Jacobian of this map has full rank at the true parameters (Rothenberg, 1971). We do not derive this rank condition analytically for this model, and we verify it numerically at the estimate. Section 7 reports the rank and condition number of the Jacobian of the simulated moments and the convergence of the estimator to the same point from several starting values.

Several restrictions are maintained throughout. The team arm raises end-of-year achievement only through the practice and the collaborations the platform records, so that the team–control contrast net of own-practice knowledge is attributed to collaboration. A unit of knowledge raises the tournament score and the end-of-year score by the same amount. This sets the scale of knowledge, and it is also a substantive restriction, since it requires the two tests to read the knowledge from practice and collaboration at the same rate. The diagnostic report is observed and was not randomized, so τ measures the response of helping to a teammate's observed report given the modeled state. Reading τ as the effect of sharing the diagnostic report among teammates requires a further assumption, that if the report were not shared, teammates would not direct extra help to those who are behind and the cost and benefits of helping would remain as in the model. School and grade shift end-of-year scores additively and by the same amount in all three arms, and every end-of-year moment is computed net of them. The structural sample conditions on an observed baseline score, and the end-of-year moments on an observed end-of-year score as well, and Appendix A examines the balance of these samples across arms.

Table 8. Targeted moments and the parameters they identify

Moment	Param.
Step 1. Tournament scores	
Score on practice, within student, both arms, collaboration controlled	θ_0
Score level	α_0
Score on baseline	α_z
Between-student share of score var.	σ_α
Score dispersion	σ_Y
Step 1. Practice and attendance	
Share with $e = 0$	κ_1
Share with $e = 1$	β^e
Mean of $\log e$, interior	ξ
$\log e$ on z_i	π_b
Within-student var. of $\log e$	σ_ν
Between-student var. of $\log e$	σ_θ
Attendance rate	μ_a
Attendance on $\log e$, within student, interior	ρ_c
Attendance on z_i	B_v
Practice on assignment length, within student	ψ
Step 2. End-of-year achievement	
Control end-of-year slope on S_i^{base}, net of school and grade	λ
Control end-of-year dispersion, net of school and grade	σ_ε
Control end-of-year level	$\delta_0 + \delta_C$
Step 3. Collaboration, helping and the control-arm level	
Individual − control, end of year	δ_C
Team − control, end of year	ζ
Help given, level	κ_h
Help given, slope on teammates shown behind	τ

Notes. Each parameter is listed with the moment that identifies it, grouped by the step of the identification argument. The moments of step 1 are computed on the individual arm, except the score on practice, which uses both platform arms with the number of collaborations as a covariate. The moments of step 2 use control students, the two end-of-year contrasts of step 3 use the platform arms and control, and the helping moments of step 3 use the team arm. The control level identifies $\delta_0 + \delta_C$, and δ_0 follows once δ_C is identified. All parameters are estimated jointly. Appendix E defines each moment and reports its fit.

7 Estimation Strategy

We estimate the twenty-two parameters jointly by the simulated method of moments (McFadden, 1989; Pakes and Pollard, 1989), matching the twenty-two moments of Table 8, which Appendix E defines. Let $\hat{m}$ denote the vector of data moments and $m_S(\Theta)$ the average of the same moments over S simulated data sets at the parameter vector Θ. The estimate minimizes

$$\left[\hat{m} - m_S(\Theta)\right]' W \left[\hat{m} - m_S(\Theta)\right], \tag{25}$$

where W is a positive definite weighting matrix. Because the number of moments equals the number of parameters, the minimizer of (25) sets the simulated moments equal to the data moments, so the estimate and its standard errors do not depend on W (Hansen, 1982; Newey and McFadden, 1994).

Each simulated data set reproduces the experiment. Every student keeps the arm, class, team, baseline score, school and grade observed in the data, and students in the platform arms also keep their assignment lengths and diagnostic reports. The simulation draws productivity and the cost, helping, attendance and score shocks. Within each class and cycle, practice, helping, attendance and the recognition probabilities are solved as the equilibrium of Section 5.7, and realized help, attendance and tournament scores are then drawn. End-of-year scores of all students follow (24). Simulated practice is recorded as a whole number of questions, as in the data. We use $S = 96$ simulation draws, held fixed across parameter vectors, and sixty Gauss–Hermite nodes for the integrals in the recognition probabilities (Appendix D.1). Every model quantity reported in the paper uses these draws and nodes.

Each moment is computed in the same way on the data and on the simulated data. Four end-of-year moments are computed net of school and grade. The control slope on the baseline score and the control dispersion are the coefficient on the baseline score and the standard deviation of the residual in a regression of control students' end-of-year scores on the baseline score and on school and grade indicators.

The individual–control and team–control contrasts come from a regression of the end-of-year scores of students in all three arms on the baseline score and on the same indicators, and the contrast for an arm is the mean residual of its students minus the mean residual of control students. The adjustment follows the design. Classes were randomized within schools, and 14 of the 22 grade-6 classes are in the team arm (Section 3.2), so a raw difference in mean scores between the team arm and control would include the difference between grade-6 students and the others. Computing the end-of-year moments net of school and grade removes this composition difference, and the school and grade effects in (24) drop out of every moment. The intent-to-treat estimates of Section 4.1 adjust for the same indicators.

Matching regression coefficients and adjusted differences computed identically on actual and simulated data is standard in simulated method of moments and is the basis of indirect inference (Gouriéroux et al., 1993; Smith, 1993).

We minimize (25) by a Gauss–Newton method with Levenberg–Marquardt damping and a line search. At each iteration, the Jacobian of the simulated moments is computed by central differences, and candidate steps along several damped Gauss–Newton directions are compared on the objective averaged over twelve points within 0.1% of each candidate, which smooths the simulation noise in the objective. A step is accepted when it lowers this average by more than two standard errors, and the search stops when no step does. Restarted from three points that perturb every parameter by 20% at random, the search returns to the same solution. At the solution, the Jacobian of the simulated moments has full rank, with condition number 260.

Standard errors use the covariance matrix $\left(1+\frac{1}{S}\right) G^{-1}\Omega\,(G^{-1})'$, where G is the Jacobian of the simulated moments at the estimate, Ω is the covariance matrix of the data moments, and the factor $1+1/S$ accounts for simulation noise (McFadden, 1989; Pakes and Pollard, 1989).

8 Empirical Estimates

This section reports the estimates, their fit to the twenty-two targeted moments and several untargeted checks. Table 9 reports the estimates and their standard errors.

Eighteen of the twenty-two parameters are distinguishable from zero at the 5% level. We discuss them by the part of the model they govern. Productivity θ_i is the number of end-of-year points a student gains by attempting every assigned question of one cycle. The median student gains $e^{\theta_0} = 2.0$ points, and the dispersion $\sigma_\theta = 1.06$ implies that a student one standard deviation above the median in log productivity gains $e^{1.06} = 2.9$ times as much from the same practice. The value of knowledge, $\beta^e = 0.29$ (0.11), measures how much a student values one point of end-of-year knowledge, in units of the cost of practice time. The mean of log productivity, $\theta_0 = 0.70$ (0.64), is imprecise on its own. Practice depends mainly on the sum $\theta_0 + \log\beta^e$, which is estimated at -0.53 (0.27), and the tournament score separates the two.

The cost of practice rises steeply with the share of questions attempted. With the curvature $\xi = 3.81$ (0.54), the marginal cost of attempting the last question of an assignment is 4.7 times the marginal cost at the halfway point. The linear cost, $\kappa_1 = 0.09$ (0.17), is the marginal cost of starting an assignment, about 8% of the marginal cost of completing it. It

Table 9. Parameter estimates

		Estimate	(SE)	p
Types and knowledge				
θ_0	mean of log productivity	0.704	(0.644)	0.274
σ_θ	dispersion of log productivity	1.064	(0.500)	0.033
ζ	knowledge per collaboration	0.441	(0.131)	< 0.001
Practice				
β^e	value of knowledge	0.291	(0.113)	0.010
ξ	cost curvature	3.812	(0.540)	< 0.001
κ_1	linear cost	0.089	(0.172)	0.606
ψ	cost elasticity to assignment length	0.422	(0.071)	< 0.001
π_b	cost loading on baseline	0.093	(0.071)	0.189
σ_ν	transitory cost dispersion	0.683	(0.221)	0.002
Attendance and recognition				
μ_a	attendance index	0.945	(0.111)	< 0.001
ρ_c	attendance loading on cost	0.327	(0.132)	0.013
B_v	value of recognition	1.447	(0.954)	0.130
Helping				
τ	value of helping a teammate shown behind	0.646	(0.284)	0.023
κ_h	cost of helping	1.171	(0.172)	< 0.001
Tournament score				
α_0	level	33.09	(0.73)	< 0.001
α_z	slope on baseline	5.44	(0.51)	< 0.001
σ_α	permanent test effect	9.59	(2.93)	0.001
σ_Y	transitory score shock	16.66	(0.36)	< 0.001
End-of-year production				
δ_0	intercept	21.33	(2.19)	< 0.001
λ	persistence of baseline	0.224	(0.033)	< 0.001
δ_C	control-arm level	10.18	(1.78)	< 0.001
σ_ε	residual dispersion	13.69	(0.38)	< 0.001

Notes. Standard errors, in parentheses, are computed as described in Section 7.

governs the share of student-cycles with no work, which the model matches exactly, and it is imprecisely estimated. The elasticity $\psi = 0.42$ (0.07) means that doubling the length of an assignment raises the cost of attempting a given share of it by 34%. The cost of time varies mostly from cycle to cycle. A one standard deviation cost shock multiplies it by $e^{\sigma_\nu} = 2.0$, and one standard deviation of the baseline score multiplies it by $e^{\pi_b} = 1.10$, an imprecisely estimated effect with $\pi_b = 0.093$ (0.071).

A student with the average baseline score, no cost shock and the average recognition probability attends a tournament with probability $\Lambda(\mu_a) = 0.72$, where $\mu_a = 0.94$ (0.11). The loading $\rho_c = 0.33$ (0.13) links attendance to the cost of time, so a one standard deviation cost shock lowers the probability of attending by about 4.5 percentage points. The value of recognition, $B_v = 1.45$ (0.95), translates the chance of recognition into attendance. A rise of 0.1 in the recognition probability raises the probability of attending by about 2.9 percentage points, an imprecisely estimated effect. In helping, $\tau = 0.65$ (0.28) multiplies the odds that

a student helps a teammate shown behind by $e^{\tau} = 1.9$, and the cost of helping, $\kappa_h = 1.17$ (0.17), sets the overall level of help. At the estimates, a student helps a given teammate who attempts the whole assignment with probability 0.26 if the teammate is not shown behind and 0.40 if the teammate is shown behind.

One collaboration adds $\zeta = 0.44$ (0.13) end-of-year points to the knowledge of each of the two students involved, about a fifth of what the median student gains from completing a cycle of practice. In the tournament score, one standard deviation of the baseline score adds $\alpha_z = 5.44$ points, and the transitory score shock has a standard deviation of $\sigma_Y = 16.7$ points, so the score of a single tournament is a noisy measure of knowledge. In the end-of-year score, $\lambda = 0.22$ (0.03) is the number of end-of-year points that one point of baseline score carries within a school and grade, and the control-arm level, $\delta_C = 10.18$ (1.78), measures what control students learn during the year away from the platform. The remaining parameters set the levels and dispersions of the two scores.

Appendix Table E.1 shows that the model fits the twenty-two targeted moments closely. The largest standardized difference is 0.15, for the within-student slope of the score on practice, and every other moment is matched within 0.04 standard errors. Appendix Figure E.1 shows that the model also reproduces the distribution of practice in the individual arm, including the masses at no work and at full completion. We also examine the fit to moments that the estimation does not target. The model reproduces the within-student association between collaboration and tournament scores and the share of attending students recognized in each arm, and Appendix E reports these and other untargeted moments.

9 Counterfactual Analysis

We use the estimated model to run three counterfactual exercises. First, we decompose the gain of the team design over the control condition into the contributions of individual platform use, organizing students in teams, and the interaction of the platform use and team collaboration (i.e., sharing the platform's diagnostic reports among teammates). Second, we trace the distributional gains from the team design across different student productivity levels. Third, we evaluate the effects of a policy that increases participation in assignments. Each exercise changes one feature of the design and solves the equilibrium of Section 5.7 again for every class and cycle at the estimates of Table 9, so that practice, helping, attendance and recognition all adjust.

9.1 Individual platform use, teams, and sharing diagnostic reports among teammates

We begin by decomposing the model's difference in end-of-year assessment scores between team and control conditions. This difference combines three effects: 1) the platform, 2) the teams, and 3) the interaction of the two (i.e., sharing the platform's diagnostic reports among teammates). The difference between individual and control classes helps to identify only the first. To separate the other two, in Counterfactual 1 we simulate a scenario that the experiment did not run, teams on a platform that does not share the diagnostic reports among teammates. In the model, sharing the reports among teammates affects behavior through τ, the value of helping a teammate shown behind, so we set $\tau = 0$ in this counterfactual and keep every other parameter at its estimate, under the assumption that the benefits of helping would remain as in the model. The help among teammates changes and practice adjusts because students anticipate the help their work attracts. The model's team–control difference is then the sum of three parts. The first is the individual–control difference. The second is the gain of Counterfactual 1 over the individual design, which measures the contribution of teams. The third is the gain of the team design over Counterfactual 1, which measures the interaction effect of team and platform (i.e., the contribution of sharing the platform's diagnostic reports among teammates).

Table 10. Individual platform use, teams, and sharing diagnostic reports among teammates

A. Differences in end-of-year scores, in points		
	Model	Data
Individual − control	1.05	1.06
Team − control	3.35	3.35
B. Decomposition of the model's team − control difference		
	Points	Share
Platform (individual − control)	1.05	31%
Teams (Counterfactual 1 − individual)	1.94	58%
Sharing diagnostic reports among teammates (team − Counterfactual 1)	0.36	11%
Total (team − control)	3.35	100%
C. Collaborations per team-class student-cycle		
Team design	1.44	
Counterfactual 1	1.19	

Notes. The differences in end-of-year scores are net of the baseline score, school and grade, as are the end-of-year moments of Section 7. Counterfactual 1 is the team design with $\tau = 0$. Parts may not add to the total because of rounding.

Table 10 reports the decomposition. Panel A shows that the model reproduces both differences in end-of-year scores in the data. Panel B shows that individual platform use

contributes 1.05 points, about 31% of the model's team–control difference, teams 1.94 points, or 58%, and sharing the platform's diagnostic reports among teammates (interaction of team and platform) 0.36 points, or 11%.

Panel C shows that compared to the team design, in Counterfactual 1 collaborations fall from 1.44 to 1.19 per team-class student-cycle, about 17%. Appendix Figure C.1 shows help received per student-cycle in Counterfactual 1 and in the team design, by whether the student is shown behind and by baseline decile. Compared to Counterfactual 1, sharing the platform's diagnostic reports among teammates in the team design directs help toward teammates shown behind, raising the help they receive from 0.59 to 0.95 per student-cycle, while help to other teammates stays at 0.60. Weaker students are shown behind more often, in 42% of student-cycles in the bottom baseline decile against 19% in the top. Help received therefore rises most at the bottom, by 0.17 per student-cycle against 0.06 in the top decile. For the same students, sharing the reports among teammates raises end-of-year achievement by 0.46 points in the bottom decile against 0.34 in the top.[11]

9.2 Who gains from the team design

Next, we examine who gains from the team design, and how its effects on practice, collaboration and end-of-year achievement vary with students' latent learning productivity and observed baseline score. In Counterfactual 2, we place each team-class student in the individual design and compare the outcome with the team design. Let's denote the individual design by I and the team design by T. The individual design has no collaboration, and the student's baseline score, school and grade effects and residual ε_i are the same under both designs, so by (24)

$$S_i^{\text{end}}(T) - S_i^{\text{end}}(I) \;=\; \theta_i \sum_t \big[e_{it}(T) - e_{it}(I)\big] + \zeta \sum_t C_{it}(T), \tag{26}$$

where $S_i^{\text{end}}(T)$ and $S_i^{\text{end}}(I)$ are student i's end-of-year scores under the two designs, $e_{it}(T)$ and $e_{it}(I)$ the student's practice, and $C_{it}(T)$ the student's collaborations under the team design. Figure 1 reports these differences by decile of productivity θ_i, and Appendix Table C.1 reports them by decile of the baseline score.

Panel (a) of Figure 1 shows that predicted end-of-year achievement rises in every productivity group. The average gain is 2.43 points, 0.15 standard deviations of the end-of-year score, and 99.5% of team-class students gain. Of this gain, 0.22 points come from the change

[11] Note that the 0.36 points in Table 10 are computed as the targeted moments are, net of baseline, school and grade.

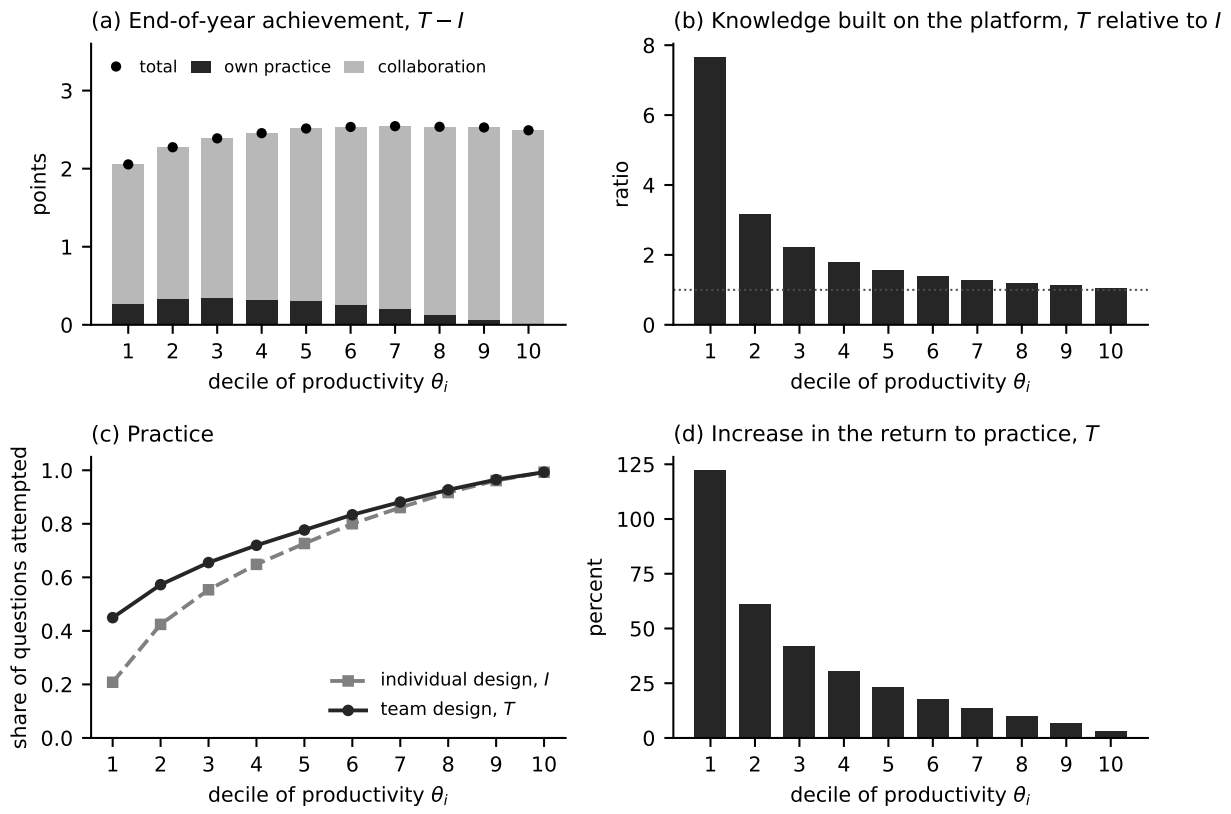


Figure 1. Who gains from the team design

Notes. The figure groups team-class students by decile of productivity θ_i. Panel (a) splits the change in end-of-year achievement into the two terms of (26). Panel (b) divides the knowledge built on the platform over the year under T by that under I. Panel (c) shows practice under I and T. Panel (d) shows the percentage increase in the marginal value of practice (20) from the help a unit of practice attracts under T, that is, ζ times help received per unit of practice divided by mean productivity in the decile.

in own practice and 2.21 from collaboration. The collaboration part rises with productivity because help is received on the questions a student attempts, and more productive students attempt more of them. The gain varies much less than productivity does, from 2.05 points in the bottom decile to a maximum of 2.54, because a collaboration produces the same knowledge, ζ, at every level of productivity. Panel (b) shows that, relative to what they learn on the platform under the individual design, the least productive students gain the most. The team design raises the knowledge that the bottom decile builds on the platform over the year from 0.31 to 2.36 points, a factor of 7.7, against a factor of 1.05 in the top decile. Appendix Table C.1 shows that the gain varies little with the baseline score. It is 2.66 points in the bottom baseline decile and between 2.33 and 2.48 in the other nine. The experimental team–individual difference in end-of-year scores compares different students, and its counterpart in the model, computed as the targeted moments are, is 2.30 points, against 2.29 in the data.

Panel (c) of Figure 1 shows that practice rises most among the least productive students. Under the individual design, the least productive tenth attempt a share 0.21 of their questions, and the team design raises this to about 0.45. The most productive tenth, at 0.99 under the individual design, changes little. Panel (a) also shows the own-practice part of the gain, which by (26) is productivity times the change in practice summed over the year. It is therefore smaller in the bottom decile, where productivity is low, close to zero in the top deciles, where practice barely changes, and largest in the third decile, at 0.34 points against 0.27 in the bottom decile. Averaged over team-class students, practice rises by 0.07, and 84% of students practice more. The marginal value of practice (20) explains the pattern. Under the team design a unit of practice also attracts help. Help received per unit of practice is 0.94 in the bottom decile and 0.92 in the top, so with $\zeta = 0.44$ the help attracted by attempting every question in a cycle is worth about 0.41 points to every student. Panel (d) shows that this raises the return to practice by 122% in the bottom decile of productivity and by 3% in the top decile. Help is directed by the diagnostic reports shared among teammates, which reflect each student's score on the base assignment. Appendix Table C.1 shows that students in the bottom baseline decile therefore receive more help, 0.87 per student-cycle against 0.62 in the top, since they are shown behind more often. The chance of recognition rises most for weaker students, by 0.34 in the bottom baseline decile against 0.09 in the top, since under the team recognition rule it depends on the team's total score. Attendance rises from 0.691 to 0.741.

9.3 Participation nudges

Finally, we examine the effectiveness of a policy that raises participation in assignments in the two platform arms. Teachers announce tournament dates in advance and do not announce assignment dates, so students typically learn of an assignment at school. Teachers and schools can raise participation by explaining the benefits of practice or by offering incentives. In the model such a policy lowers κ_1, the linear part of the cost of practice. By (23), a student does no work when the marginal value of practice per unit of cost is at most κ_1, so a lower κ_1 brings more students to start and raises the practice of those already working, and in team classes the extra work also attracts help. In Counterfactual 3, we lower κ_1 in steps, $\kappa_1(s) = s\,\hat{\kappa}_1$ for $s \in \{1, 0.75, 0.5, 0.25, 0\}$, where $\hat{\kappa}_1$ is the estimate, solve both arms again with each class's own recognition rule, and report for each arm $r \in \{\text{team}, \text{individual}\}$ the gain

$$G^r(s) = \mathbb{E}\big[S_i^{\text{end}} \mid r,\ \kappa_1(s)\big] - \mathbb{E}\big[S_i^{\text{end}} \mid r,\ \hat{\kappa}_1\big], \tag{27}$$

where the expectation is over the arm's students and the draws.

Table 11. A participation nudge, by arm

	Any work		Gain in end-of-year points				Collaborations
s	Team	Individual	Team	of which collaboration	Individual	Difference	Team
1.00	0.979	0.926	0.00	0.00	0.00	0.00	1.44
0.75	0.990	0.952	+0.10	+0.04	+0.08	+0.02	1.46
0.50	0.997	0.977	+0.19	+0.07	+0.15	+0.04	1.49
0.25	1.000	0.994	+0.28	+0.10	+0.22	+0.06	1.51
0.00	1.000	1.000	+0.35	+0.13	+0.28	+0.08	1.53

Notes. The gains are $G^r(s)$ of (27). Any work is the share of student-cycles with $e > 0$, and collaborations are per student-cycle in team classes.

Table 11 reports the results. Removing the linear cost brings every student-cycle to some work and raises end-of-year achievement by 0.35 points in team classes against 0.28 in individual classes, 28% more. The difference grows at every step, $0.02, 0.04, 0.06, 0.08$ points as κ_1 falls to three quarters, half, a quarter and zero of its estimate. In team classes, 0.22 of the 0.35 points come from own practice and 0.13 from collaboration, as collaborations per student-cycle rise from 1.44 to 1.53. Appendix Figure C.2 shows the gains at $\kappa_1 = 0$ by decile of productivity. Of the extra gain in team classes, 60% accrues to the less productive half of students, who work least at the estimated cost.

10 Conclusion and Implications

This paper asks whether organizing students in teams makes educational technology more effective, and why. To answer this question, we designed and built an EdTech platform for grade 3–6 mathematics that diagnoses each student's knowledge gaps, assigns personalized practice and runs in-class tournaments with classroom recognition. We implemented the platform in two designs. In the individual design, each student sees only their own diagnostic reports and competes individually. In the team design, students work in ability-balanced teams, see their teammates' diagnostic reports and compete as teams. In partnership with a non-governmental organization, we ran a pre-registered randomized field experiment with 3,294 students in 92 classes of 13 schools in 6 cities and 3 Indian states over two academic years. During the experiment, the platform delivered more than 34 thousand practice assignments and recorded more than 0.4 million item responses. Because we built the platform ourselves, we could implement both designs exactly and record practice, help within teams and tournament participation for every student and cycle. These records make it possible to study how the design of a platform shapes the way students practice and learn from one another.

The experiment shows that the team design raises end-of-year mathematics achievement

by 3.8 points, or 0.23 standard deviations, relative to the control condition without the platform. The effect of individual platform use is positive but imprecise, as is the team–individual difference of 2.4 points. Average practice and tournament attendance are similar in the two platform arms. Within teams, students direct more help toward teammates whose diagnostic reports show learning gaps, and collaboration is associated with higher tournament scores within students.

These patterns guide a structural model of practice, helping within teams, tournament attendance and skill accumulation, which we estimate from the experimental data and use to evaluate designs that the experiment did not implement. The model attributes 31% of the team design's gain over the control condition to individual platform use, 58% to organizing students in teams, and 11% to the interaction of the platform and teams, that is, to sharing the platform's diagnostic reports among teammates. If the reports were not shared among teammates, collaboration would fall by about 17%, and teammates shown behind would receive no more help than other teammates.

The simulations also show who gains from the team design. Predicted achievement rises in every productivity group, by between 2.05 and 2.54 points, because a collaboration produces the same knowledge at every level of productivity. Relative to what they learn on the platform under the individual design, the least productive students gain the most, because the help their practice attracts raises their return to practice the most. Removing the linear cost of practice raises predicted achievement by 0.35 points in team classes and 0.28 points in individual classes, which points to a complementarity between participation and collaboration.

These gains come at a modest cost. The platform used 200 minutes of class time per month and cost INR 500, approximately USD 6, per participating student per month, including software development, infrastructure, and staffing.[12] The team design required no more resources than the individual design. As artificial intelligence makes personalized content cheaper to produce, our results suggest that platforms can gain further by organizing how classmates learn from one another, and that policies that raise participation in practice yield more when such collaboration is in place.

The model treats the diagnostic reports as given when practice changes, applies a common return to collaboration across students, and treats the cycles of the year as separable. Experiments that vary, within team classes, whether teammates see each other's diagnostic reports would test the effect of sharing the reports directly. More broadly, the organization of peer interactions deserves a central place in the design and evaluation of classroom

[12]The scheduled monthly allocation is two 40-minute base assignments, two 40-minute personalized assignments, and one 40-minute tournament. Actual exposure varies across the recorded cycles.

technology.

References

Maryam Alavi and Dorothy E. Leidner. Research commentary: Technology-mediated learning—a call for greater depth and breadth of research. *Information Systems Research*, 12(1):1–10, 2001.

Orazio P. Attanasio, Costas Meghir, and Ana Santiago. Education choices in Mexico: Using a structural model and a randomized experiment to evaluate PROGRESA. *Review of Economic Studies*, 79(1):37–66, 2012.

Abhijit Banerjee, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. Mainstreaming an effective intervention: Evidence from randomized evaluations of "teaching at the right level" in india. NBER Working Paper 22746, National Bureau of Economic Research, 2016.

Abhijit V. Banerjee, Shawn Cole, Esther Duflo, and Leigh Linden. Remedying education: Evidence from two randomized experiments in india. *Quarterly Journal of Economics*, 122(3):1235–1264, 2007.

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

Nicholas Bloom, Raffaella Sadun, and John Van Reenen. Americans do IT better: US multinationals and the productivity miracle. *American Economic Review*, 102(1):167–201, 2012.

Timothy F. Bresnahan, Erik Brynjolfsson, and Lorin M. Hitt. Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quarterly Journal of Economics*, 117(1):339–376, 2002.

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

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

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

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

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

Gary Chamberlain. Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225–238, 1980.

Christopher S. Cotton, Brent R. Hickman, John A. List, Joseph Price, and Sutanuka Roy. Why don't struggling students do their homework? disentangling motivation and study productivity as drivers of human capital formation. *Journal of Political Economy*, 134(1): 86–149, 2026.

Esther Duflo, Rema Hanna, and Stephen P. Ryan. Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–1278, 2012.

Dennis Epple and Richard E. Romano. Peer effects in education: A survey of the theory and evidence. In *Handbook of Social Economics*, volume 1, pages 1053–1163. 2011.

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

Roland G. Fryer. Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics*, 126(4):1755–1798, 2011.

Christian Gouriéroux, Alain Monfort, and Eric Renault. Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118, 1993.

Saurabh Gupta and Robert P. Bostrom. Technology-mediated learning: A comprehensive theoretical model. *Journal of the Association for Information Systems*, 10(9):686–714, 2009.

Juho Hamari, Jonna Koivisto, and Harri Sarsa. Does gamification work?—a literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)*, pages 3025–3034, 2014.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

Caroline M. Hoxby and Gretchen Weingarth. Taking race out of the equation: School reassignment and the structure of peer effects. Working paper, Harvard University, 2005.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York, 2015.

Anuj Kumar and Anurag Garg. Improving skill production with peer-induced knowledge diffusion in k-12 schools. Working Paper 5145197, SSRN, 2025.

Victor Lavy, Olmo Silva, and Felix Weinhardt. The good, the bad and the average: Evidence on the scale and nature of ability peer effects in schools. NBER Working Paper 15600, National Bureau of Economic Research, 2009.

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

Steven D. Levitt, John A. List, Susanne Neckermann, and Sally Sadoff. The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4):183–219, 2016.

Winston Lin. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics*, 7(1):295–318, 2013.

De Liu, Radhika Santhanam, and Jane Webster. Toward meaningful engagement: A frame-

work for design and research of gamified information systems. *MIS Quarterly*, 41(4): 1011–1034, 2017.

Daniel McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026, 1989.

Karthik Muralidharan and Abhijeet Singh. Adapting for scale: Experimental evidence on technology-aided instruction in india. NBER Working Paper 34205, National Bureau of Economic Research, 2025.

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

Karthik Muralidharan, Mauricio Romero, and Kaspar Wüthrich. Factorial designs, model selection, and (incorrect) inference in randomized experiments. *Review of Economics and Statistics*, 107(3):589–604, 2025.

Whitney K. Newey and Daniel McFadden. Large sample estimation and hypothesis testing. In Robert F. Engle and Daniel L. McFadden, editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier, Amsterdam, 1994.

Ariel Pakes and David Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5):1027–1057, 1989.

Thomas J. Rothenberg. Identification in parametric models. *Econometrica*, 39(3):577–591, 1971.

Bruce Sacerdote. Peer effects with random assignment: Results for dartmouth roommates. *Quarterly Journal of Economics*, 116(2):681–704, 2001.

Bruce Sacerdote. 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*, volume 3, pages 249–277. 2011.

Anthony A. Smith. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8(S1):S63–S84, 1993.

Petra E. Todd and Kenneth I. Wolpin. Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review*, 96(5):1384–1417, 2006.

Jacob L. Vigdor and Thomas Nechyba. Peer effects in North Carolina public schools. In Ludger Woessmann and Paul E. Peterson, editors, *Schools and the Equity Tradeoff*. MIT Press, Cambridge, MA, 2007.

Markus Weinmann, Christoph Schneider, and Jan vom Brocke. Digital nudging. *Business & Information Systems Engineering*, 58(6):433–436, 2016.

Shlomo Yitzhaki. On using linear regressions in welfare economics. *Journal of Business & Economic Statistics*, 14(4):478–486, 1996.

Alwyn Young. Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics*, 134(2): 557–598, 2019.

APPENDIX

A Sample balance and attrition checks

This appendix checks that random assignment produced comparable arms, that missing endline scores leave the arms comparable, and that the intent-to-treat estimates of Section 4.1 are robust to the choice of covariates and estimator. Classes were randomly assigned to the three arms (*control*, *individual*, *team*) separately within each school. The randomization sample contains 3,294 students in 92 classes across 13 schools and four grades (3–6).

Because the class is the unit of assignment, every statistical test in this appendix either collapses the data to the class or clusters standard errors by class, and randomization inference reassigns arm labels across classes within each school.

A.1 Number of classes in each arm

This subsection checks whether the number of classes is balanced across arms, overall and within each grade. Table A.1 reports the number of classes assigned to each arm in each grade. A χ^2 test of whether each arm has a third of the classes gives $\chi^2(2) = 1.07$ and $p = 0.587$, so the numbers of classes are consistent with an equal split across the three arms. The numbers are similar across arms in grades 3 and 4, and the team arm has few grade-5 classes and many grade-6 classes, which Section 3.2 discusses. A randomization-inference test of independence between arm and grade, with 20,000 within-school permutations, gives $p = 0.134$.

Table A.1. Distribution of classes across arms, by grade

	Number of classes				Row share (%)		
Grade	Control	Individual	Team	Total	Control	Individual	Team
3	8	8	8	24	33.3	33.3	33.3
4	8	7	9	24	33.3	29.2	37.5
5	8	10	4	22	36.4	45.5	18.2
6	3	5	14	22	13.6	22.7	63.6
Total	27	30	35	92	29.3	32.6	38.0

Notes. Row shares are percentages of each grade's classes.

A.2 Balance of baseline scores and demographic characteristics

This subsection checks whether baseline scores and demographic characteristics are balanced across arms. Table A.2 compares the mean baseline score and the demographic characteristics of students across the three arms.

Section 3.2 discusses the baseline score, the female share and the availability of a baseline score. The other characteristics, caste, parental and sibling education and family income,

Table A.2. Balance of baseline scores and demographic characteristics across arms

	Mean			Normalized diff.		
Variable	Control	Individual	Team	Ind.−Ctl	Team−Ctl	RI p-value
Cognitive						
Baseline score	42.06	41.90	41.34	−0.01	−0.04	0.923
Baseline score missing	0.14	0.09	0.06	−0.18	−0.26	0.040
Demographics						
Female	0.31	0.57	0.51	+0.54	+0.41	0.107
High caste (general)	0.93	0.90	0.93	−0.11	+0.00	0.740
Father's education < 8th	0.49	0.48	0.49	−0.02	+0.01	0.952
Mother's education < 8th	0.48	0.44	0.46	−0.08	−0.04	0.799
Sibling education < 8th	0.61	0.54	0.54	−0.13	−0.13	0.143
Income < INR 25,000	0.99	0.99	0.98	+0.04	−0.06	0.621

Notes. A normalized difference is the difference between the arm mean and the control mean divided by the square root of the average of the two variances (Imbens and Rubin, 2015). Each row is a regression of the characteristic on the two arm indicators with school fixed effects, and the RI p-value is the randomization-inference p-value of the joint test that both arm coefficients are zero, computed from 5,000 within-school permutations of the arm labels.

are balanced across arms.

A.3 Missing endline scores

This subsection checks whether the students with an endline score remain comparable across arms. Section 3.2 shows that the endline response rate does not differ significantly across arms.

Table A.3 compares student characteristics across the three arms in two samples. Panel A uses all randomized students, and Panel B drops the students without an endline score. Comparing the two panels shows whether the students who miss the endline assessment change the balance across arms. The table reports the normalized difference for each pair of arms. Among students with an endline score, the demographic characteristics that Section 3.2 does not discuss remain balanced across arms.

A.4 Robustness of the intent-to-treat estimates

Table A.4 examines the robustness of the intent-to-treat estimates of Table 3 to the covariates and the estimator. Each column is one regression on the students of all three arms. Control classes are the comparison group, so the coefficients on the individual and team indicators are the individual–control and team–control effects, and the table also reports their difference. Column (1) has no covariates and no fixed effects. Column (2) adds the baseline score, the missing-baseline indicator and school and grade fixed effects, and column (3) adds the female indicator, which gives the specification of Table 3. Column (4) replaces the school and grade fixed effects with school-by-grade fixed effects, and column (5) applies the fully interacted estimator of Lin (2013), which demeans the baseline score, the missing-baseline indicator

Table A.3. Covariate balance in the randomized sample and among students with an endline score

	Mean (observations)			Normalized difference		
	Control	Individual	Team	Ind.−Ctl	Team−Ctl	Team−Ind.
A. Randomized sample						
Students	936	1,122	1,236			
Baseline score	42.06 (802)	41.90 (1,026)	41.34 (1,157)	−0.01	−0.04	−0.03
Baseline missing	0.14 (936)	0.09 (1,122)	0.06 (1,236)	−0.18	−0.26	−0.08
Female	0.31 (936)	0.57 (1,122)	0.51 (1,236)	+0.54	+0.41	−0.12
Grade 3	0.30 (936)	0.29 (1,122)	0.25 (1,236)	−0.01	−0.10	−0.09
Grade 4	0.33 (936)	0.24 (1,122)	0.29 (1,236)	−0.21	−0.09	+0.12
Grade 5	0.25 (936)	0.31 (1,122)	0.12 (1,236)	+0.12	−0.35	−0.47
Grade 6	0.12 (936)	0.16 (1,122)	0.34 (1,236)	+0.12	+0.54	+0.42
General caste	0.93 (635)	0.90 (804)	0.93 (864)	−0.11	0.00	+0.11
Father's education < 8th	0.49 (630)	0.48 (794)	0.49 (857)	−0.02	+0.01	+0.03
Mother's education < 8th	0.48 (630)	0.44 (794)	0.46 (857)	−0.08	−0.04	+0.05
Sibling education < 8th	0.61 (619)	0.54 (780)	0.54 (834)	−0.13	−0.13	0.00
Income < INR 25,000	0.99 (630)	0.99 (794)	0.98 (857)	+0.04	−0.06	−0.10
B. Endline respondents						
Students	841	955	1,125			
Baseline score	42.35 (738)	42.21 (889)	41.45 (1,060)	−0.01	−0.05	−0.04
Baseline missing	0.12 (841)	0.07 (955)	0.06 (1,125)	−0.18	−0.23	−0.05
Female	0.31 (841)	0.55 (955)	0.52 (1,125)	+0.51	+0.44	−0.06
Grade 3	0.30 (841)	0.32 (955)	0.26 (1,125)	+0.05	−0.07	−0.12
Grade 4	0.34 (841)	0.22 (955)	0.30 (1,125)	−0.28	−0.10	+0.17
Grade 5	0.24 (841)	0.30 (955)	0.12 (1,125)	+0.14	−0.32	−0.46
Grade 6	0.12 (841)	0.16 (955)	0.32 (1,125)	+0.11	+0.50	+0.38
General caste	0.93 (597)	0.89 (686)	0.93 (810)	−0.15	−0.01	+0.15
Father's education < 8th	0.49 (594)	0.48 (677)	0.49 (803)	−0.03	0.00	+0.03
Mother's education < 8th	0.48 (594)	0.45 (677)	0.46 (803)	−0.06	−0.04	+0.02
Sibling education < 8th	0.61 (584)	0.54 (663)	0.54 (781)	−0.14	−0.14	0.00
Income < INR 25,000	0.99 (594)	0.99 (677)	0.98 (803)	+0.05	−0.06	−0.10

Notes. Panel A includes all 3,294 randomized students in 92 classes. Panel B includes the 2,921 students with an endline score in 91 classes. Parentheses report the number of students with an observed value for each characteristic in each arm. All variables except baseline score are indicators, so their means are shares. Normalized differences are the first arm mean minus the second, divided by the square root of the average of their sample variances.

and the female indicator and interacts each with both arm indicators. Stars are based on randomization-inference p-values. The cluster-robust and wild cluster bootstrap p-values, which Table 3 reports for column (3), lead to the same conclusions for the individual–control and team–control effects.

Table A.4. Robustness of the intent-to-treat effects

	(1)	(2)	(3)	(4)	(5)
Individual − control	1.018	1.038	1.414	1.519	0.903
	(2.047)	(1.313)	(1.277)	(1.636)	(1.155)
Team − control	2.089	3.440**	3.764**	4.266*	3.272**
	(2.141)	(1.333)	(1.325)	(1.664)	(1.129)
Team − individual	1.071	2.402	2.350	2.747	2.369
	(2.105)	(1.370)	(1.313)	(1.471)	(1.253)
Baseline score and missing-baseline indicator	–	Yes	Yes	Yes	Yes
Female indicator	–	–	Yes	Yes	Yes
School and grade fixed effects	–	Yes	Yes	–	Yes
School-by-grade fixed effects	–	–	–	Yes	–
Covariates interacted with arm indicators	–	–	–	–	Yes
Students	2,921	2,921	2,921	2,921	2,921
Classes	91	91	91	91	91

Notes. Each column is a regression of the endline score, in points (SD 16.07), on the individual and team indicators and the listed covariates, estimated on the students of all three arms with control classes as the comparison group. The team − individual row is the difference between the two coefficients of the same regression. Column (3) is the specification of Table 3. Standard errors, clustered by class, are in parentheses. Stars are based on randomization-inference p-values from 2,000 permutations of arm labels across classes within school, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

B Additional reduced-form evidence

This appendix reports the reduced-form estimates that Section 4 summarizes and the descriptive evidence on practice, attendance and recognition that Section 5 draws on. Table B.1 asks whether the association between the platform's reports and help in Section 4.2 comes from differences across students or from changes over cycles for the same student. Column (1) pools all student-cycles, and column (2) adds student fixed effects, which compare cycles of the same student. Section 4.2 discusses the result.

Table B.1. Diagnostic reports and help within students

	Pooled	Student fixed effects
Help given, behind-minus-ahead coefficient	0.114***	0.002
	(0.029)	(0.025)
Help received, student shown behind	0.076***	0.013
	(0.025)	(0.023)
Observations, help given	3,891	3,891
Observations, help received	3,262	3,262

Notes. Both specifications include class-by-cycle fixed effects, the number of teammates working and an indicator for whether the student worked on the assignment. The pooled specification also includes the student's baseline score relative to teammates as a covariate. The help-received regressions also include the student's practice on the whole assignment and on the personalized assignment. The help-given row reports the difference between the coefficients on teammates shown behind and ahead. Standard errors, clustered by class, are in parentheses. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

B.1 Additional tournament specifications

This subsection checks how the association between collaboration and tournament scores of Section 4.3 depends on the specification. Table B.2 reports the two specifications of Section 4.3 that extend the team-arm regressions of Table 5. Column (1) adds the interaction between practice and collaboration, and column (2) separates help received from help given.

Table B.2. Interactions and directions of help in the team arm

	(1)	(2)
Practice e_{it}	1.384	0.817
	(1.845)	(1.822)
Teammates working F_{it}	0.305	0.319
	(0.802)	(0.798)
Collaboration C_{it}	1.064*	
	(0.556)	
$e_{it} \times C_{it}$	−0.586	
	(0.637)	
Help received H_{it}		0.646
		(0.441)
Help given O_{it}		0.511*
		(0.286)
p-value, received = given		0.824
Student fixed effects	Yes	Yes
Class-by-tournament fixed effects	Yes	Yes
Observations	2,719	2,719

Notes. Both columns use attending students in the team arm and extend (4). Standard errors, clustered by class, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.3 estimates the tournament regression (4) for attending students in both platform arms. Column (1) includes the practice and the collaboration of the current cycle. Column (2) adds the practice and the collaboration of earlier cycles, which Section 5.2 discusses. Column (3) asks whether the association between practice and the score varies with the baseline score. The interaction, 0.33 (1.25), is not statistically significant, consistent with the model, in which productivity does not depend on the baseline score.

B.2 Practice and assignment length

This subsection documents how practice varies with the length of the assignment and how it is distributed between its two bounds. The number of questions assigned in a cycle ranges from 15 to 255, with a mean of 67 and a standard deviation of 40. Between the tenth and

Table B.3. Additional robustness checks for tournament performance

	(1)	(2)	(3)
Practice, e_{it}	1.848	1.522	1.850
	(1.133)	(1.542)	(1.136)
Collaboration, C_{it}	0.559***	0.792***	0.561***
	(0.193)	(0.271)	(0.192)
Practice in earlier cycles, $\sum_{s<t} e_{is}$		−0.651	
		(1.300)	
Collaboration in earlier cycles, $\sum_{s<t} C_{is}$		0.347	
		(0.245)	
$e_{it} \times z_i$			0.329
			(1.252)
Student fixed effects	Yes	Yes	Yes
Class-by-tournament fixed effects	Yes	Yes	Yes
Observations	5,013	5,013	5,013

Notes. The table estimates (4) for attending students in both platform arms. Standard errors, clustered by class, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

ninetieth percentiles, it runs from 20 to 135. We estimate three regressions,

$$e_{it} = b_N \log N_{it} + \phi_i + \varepsilon_{it}, \tag{B.1}$$

$$e_{it} = b_N \log N_{it} + b_{Nz}\, z_i \log N_{it} + \phi_i + \varepsilon_{it}, \tag{B.2}$$

$$\log N^p_{it} = c_e\, e^b_{it} + c_z\, z_i + \phi_{c(i)t} + \varepsilon_{it}, \tag{B.3}$$

where e_{it} is the share of the cycle's assigned questions that student i attempts in cycle t, N_{it} counts the questions in the cycle's base and personalized assignments together, z_i is the standardized baseline score, N^p_{it} is the length of the personalized assignment, e^b_{it} is the share of the base assignment attempted, and ϕ_i and $\phi_{c(i)t}$ are student and class-by-cycle fixed effects as in (4). Table B.4 reports one regression per column. Columns (1), (3) and (4) estimate (B.1) with e_{it}, $\mathbf{1}\{e_{it} = 1\}$ and attendance a_{it} as the outcome, column (2) estimates (B.2) and column (5) (B.3).

Column (1) of Table B.4 reports the within-student slope of practice on assignment length that Section 5 uses. Column (2) shows that the effect of length does not vary with the baseline score, with an interaction of −0.010 (0.009), so length affects strong and weak students alike. Column (3) shows that the effect is concentrated at full completion ($e_{it} = 1$), whose probability falls by 0.35 (0.02) per log point. Column (4) shows that attendance does not vary with length, 0.005 (0.020), since the tournament is a fixed event regardless of the cycle's assignments. Column (5) reports the regression behind the second of the two facts on how the length of the assignment is set that Section 6 uses.

Figure B.1 shows how practice varies with the length of the assignment. Panel (a) shows that the share of student-cycles completing everything declines steeply as assignments lengthen. This share is 0.501 in the shortest third of student-cycles, whose median length is 30 questions, 0.331 in the middle third, with a median of 60, and 0.138 in the longest third,

Table B.4. Practice and assignment length

	(1) e_{it}	(2) e_{it}	(3) $\mathbf{1}\{e_{it}=1\}$	(4) a_{it}	(5) $\log N^p_{it}$
$\log N_{it}$, b_N	-0.071^{***}	-0.070^{***}	-0.346^{***}	0.005	
	(0.010)	(0.010)	(0.019)	(0.020)	
$z_i \log N_{it}$, b_{Nz}		-0.010			
		(0.009)			
Base share attempted, e^b_{it}, c_e					0.013^{**}
					(0.006)
Baseline, z_i, c_z					0.000
					(0.001)
Student fixed effects	Yes	Yes	Yes	Yes	–
Class-by-cycle fixed effects	–	–	–	–	Yes
Observations	7,372	7,372	7,372	7,372	7,111

Notes. Standard errors, clustered by class, are in parentheses. $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.10$.

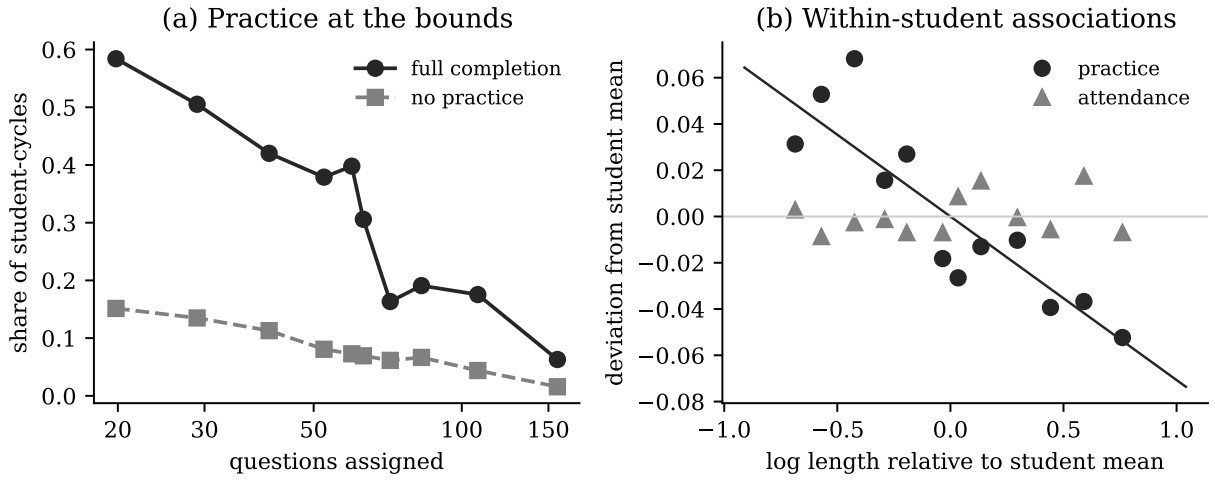


Figure B.1. Practice and the length of the assignment

Notes. Panel (a) groups student-cycles into ten bins of assignment length and shows the share completing everything and the share doing nothing in each. Panel (b) expresses each variable as a deviation from its student mean and groups these deviations into twelve bins of length. The line in panel (b) is the within-student slope of −0.071 in column (1) of Table B.4.

with a median of 100. The share doing nothing changes much less, from 0.13 in the shortest third to 0.05 in the longest. Panel (b) removes each student's own mean and shows that the relation holds within student.

Table B.5 reports the distribution of practice across student-cycles in Panel A and across students in Panel B. Practice is a share of questions attempted, so it lies between zero and one, and Panel A shows that many student-cycles are at each bound. Because practice is recorded as a whole number of questions, interior values lie on a grid set by the length of the assignment, and the estimation records simulated practice in the same way.

Panel B shows that students move in and out of the bounds from cycle to cycle. Only

Table B.5. The distribution of practice

	Share
*A. Student-cycles (*7,372*)*	
$e_{it} = 0$	0.083
$0 < e_{it} < 1$	0.589
$e_{it} = 1$	0.328
*B. Students (*2,144*)*	
Mean of e_{it} over cycles ≥ 0.95	0.106
$e_{it} = 1$ in every cycle	0.036
Mean of e_{it} over cycles ≤ 0.05	0.029

3.6% of students complete every assignment in every cycle, although 32.8% of student-cycles are at full completion. Only 2.9% of students attempt on average 5% of their questions or less, although 8.3% of student-cycles are at zero. As Table B.4 shows, the length of the assignment is a large part of what decides when students complete everything.

B.3 Attendance, practice, help and recognition

This subsection documents further patterns in the data that the model of Section 5 draws on. They concern how practice and attendance move together, how practice, tournament scores and attendance vary with the baseline score, whether recognition in one tournament is associated with practice and attendance in the next, how the help a student receives varies with practice, and how much of the variation in practice, attendance and tournament scores occurs within students. Four regressions describe the first four patterns,

$$a_{it} = b_{ae}\, e_{it} + \phi_i + \phi_{c(i)t} + \varepsilon_{it}, \tag{B.4}$$

$$W_{it} = b_z\, z_i + \phi_{c(i)t} + \varepsilon_{it}, \tag{B.5}$$

$$W_{it} = b_R\, R_{i,t-1} + b_z\, z_i + \phi_{c(i)t} + \varepsilon_{it}, \tag{B.6}$$

$$H_{it} = b_{He}\, e_{it} + \phi_{c(i)t} + \varepsilon_{it}, \tag{B.7}$$

where a_{it} equals one if student i attends the tournament of cycle t, e_{it} is the share of the cycle's questions the student attempts, z_i is the standardized baseline score, and ϕ_i and $\phi_{c(i)t}$ are student and class-by-cycle fixed effects. In (B.5), the outcome W_{it} is the log of practice in cycles with $0 < e_{it} < 1$, the tournament score Y_{it} or attendance a_{it}. In (B.6), W_{it} is either practice or attendance, $R_{i,t-1}$ equals one if the student was recognized at the previous tournament, and the sample is the students who attended that tournament. In (B.7), H_{it} is the number of times the student receives help in the cycle, in team classes, as in Section 4.3. Table B.6 reports the estimates of these four regressions.

Column (1) shows that within a student, a cycle with more practice is a cycle with higher attendance, by 0.146 (0.024) per unit of practice. Column (2) shows that attendance rises with the baseline score, by 0.028 (0.008) per standard deviation of the baseline score. Estimated separately by arm, this slope is 0.038 (0.009) in team classes and 0.016 (0.015) in individual classes. Columns (3) and (4) show that interior log practice and tournament

Table B.6. Attendance, practice, baseline and last cycle's recognition

	(1) a_{it}	(2) a_{it}	(3) $\log e_{it}$	(4) Y_{it}	(5) e_{it}	(6) a_{it}	(7) H_{it}
Practice, e_{it}	0.146***						0.909***
	(0.024)						(0.120)
Baseline, z_i		0.028***	0.032***	3.078***	0.022***	0.019*	
		(0.008)	(0.009)	(0.438)	(0.007)	(0.010)	
Recognized last cycle, $R_{i,t-1}$					0.009	−0.002	
					(0.007)	(0.018)	
Student fixed effects	Yes	–	–	–	–	–	–
Class-by-cycle fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,372	7,372	4,343	5,013	3,491	3,491	3,891

Notes. Column (1) estimates (B.4), columns (2)–(4) (B.5), columns (5)–(6) (B.6) and column (7) (B.7). Standard errors, clustered by class, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

scores also rise with the baseline score. Columns (5) and (6) show that among students who attended the previous tournament, recognition has no statistically significant association with practice or attendance in the next cycle. Column (7) shows that in team classes the help a student receives rises with the share of the assignment the student attempts, by 0.909 (0.120).

Table B.7 splits the variance of practice and attendance into two parts. The first column is the between-student part, the variance of student means, the second column is the within-student part, the variance of each student's deviations from their own mean, and the third column is the within-student part as a share of the sum of the two. The within-student share is 0.36 for log practice in cycles with $0 < e_{it} < 1$, 0.41 for practice over all cycles and 0.54 for attendance, so both decisions vary substantially from cycle to cycle for the same student. The within-student deviations of practice and attendance have a correlation of 0.095. For tournament scores, which the table does not include, the variance of student means is 58% of the total variance.

Table B.7. Permanent and transitory variation in practice and attendance

	Between students	Within student	Within share	Student-cycles	Students
Log practice, cycles with $0 < e_{it} < 1$	0.129	0.074	0.36	4,343	2,004
Practice e_{it}, all cycles	0.057	0.039	0.41	7,372	2,144
Attendance a_{it}	0.102	0.120	0.54	7,372	2,144

Notes. The table covers platform students with a baseline score. The between-students column is the variance of student means, the within-student column is the variance of deviations from student means, and the within share is the second over their sum.

B.4 What the recognition rule awards

This subsection checks that the recorded recognition outcomes follow the two recognition rules and documents the ties that the scoring produces. In individual classes the protocol recognizes the five highest tournament scores in a class-cycle; in team classes it recognizes every attending member of the three teams with the highest total score of attending members.

Table B.8 reports the evidence, which Section 5.2 summarizes, for the individual recognition rule in Panel A and for the team recognition rule in Panel B. Panel A shows that under the individual recognition rule the recognized set is exactly the top of the score ranking in 91 of 92 class-cycles.

Ties arise from the scoring. The score is the share of questions answered correctly, in points out of 100, so in a tournament of Q_{ct} questions scores move in steps of $100/Q_{ct}$. Tournaments have between 10 and 32 questions, and most of them are short. Nine in ten have between 10 and 20 questions, so in most tournaments each question moves the score by 5 points or more. Panel A of Table B.8 reports the step, which we read off each class-cycle's recorded scores. The step is 6.67 points in 76 of the 200 class-cycles in the two platform arms, 5 points in 56 and 10 points in 44. Panel A also shows that in a class-cycle with 27 attending students, the recorded scores take only 7.7 distinct values on average, so several students typically share the fifth-highest score, and all of them are recognized. The number of students scoring at or above the fifth-highest score exceeds five by 2.32 per class-cycle on average, and the number recognized exceeds five by 2.31. The number recognized under the individual recognition rule is five in a quarter of class-cycles and seven or more in half. Panel B shows that under the team recognition rule the top three teams by total score coincide with the recognized teams in 89% of class-cycles, and the remaining class-cycles have ties for third place.

Table B.8. What the recognition rule awards

A. Individual recognition of the top five scores	
Students recognized per class-cycle, mean	7.2
Students attending per class-cycle, mean	26.9
Distinct scores, class-cycle of 27 attending	7.7
Class-cycles by score step, 6.67 / 5 / 10 / other	76 / 56 / 44 / 24
At or above the fifth score, beyond five	2.32
Recognized beyond five	2.31
Recognized set is the top of the ranking	91 of 92
B. Team recognition of the top three total scores	
Spearman correlation, recorded rank and rank by total score	0.996
Top three teams by total score are the recognized teams	89%

Notes. Values are means per class-cycle where stated. Panel A covers the individual classes, except the score-step row, which covers the 200 class-cycles of both arms.

C Additional counterfactual results

This appendix reports additional results for the counterfactuals of Section 9. Figure C.1 shows how sharing the platform's diagnostic reports among teammates changes who receives help, which Section 9.1 discusses. Table C.1 reports Counterfactual 2 by decile of the baseline score, and Figure C.2 shows the gains from the participation nudge by decile of productivity, which Sections 9.2 and 9.3 discuss.

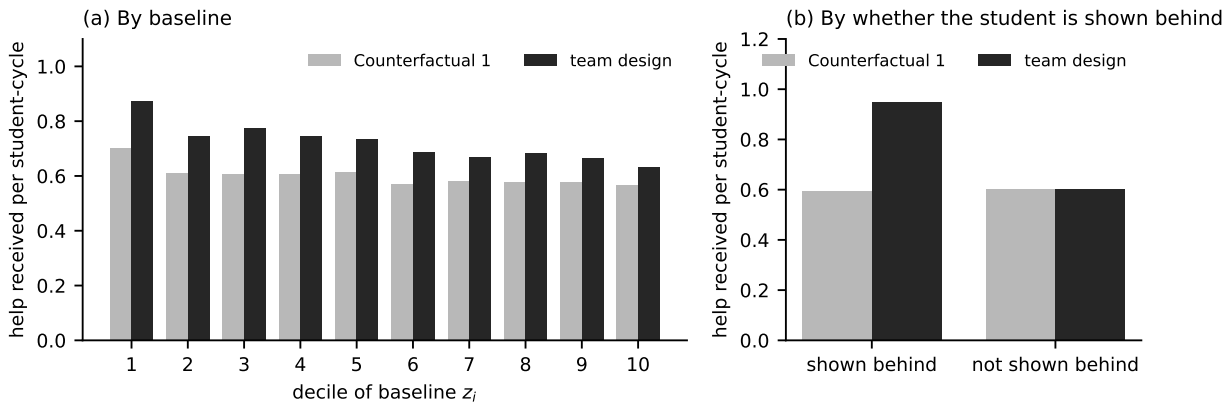


Figure C.1. Diagnostic reports shared among teammates and who receives help

Notes. The figure shows help received per student-cycle in team classes, in Counterfactual 1, in which the diagnostic reports are not shared among teammates, and in the team design. Panel (a) groups students by decile of the baseline score, and panel (b) by whether the student is shown behind.

Table C.1. Who gains, by baseline score

	Practice		Collaborations	Help received	Change in end-of-year points, $T - I$		
	I	T	T	T	Own practice	Collaboration	Total
By decile of the baseline score z_i							
Decile 1	0.739	0.814	1.66	0.87	0.24	2.42	2.66
Decile 2	0.721	0.793	1.47	0.75	0.23	2.15	2.38
Decile 3	0.715	0.788	1.48	0.78	0.23	2.18	2.41
Decile 4	0.717	0.784	1.44	0.74	0.23	2.18	2.41
Decile 5	0.704	0.773	1.45	0.72	0.22	2.13	2.35
Decile 6	0.703	0.770	1.37	0.69	0.22	2.11	2.33
Decile 7	0.706	0.772	1.38	0.66	0.22	2.23	2.46
Decile 8	0.708	0.774	1.41	0.69	0.23	2.25	2.48
Decile 9	0.699	0.763	1.37	0.67	0.22	2.18	2.40
Decile 10	0.682	0.743	1.34	0.62	0.21	2.24	2.45
All team-class students	0.709	0.777	1.44	0.72	0.22	2.21	2.43

Notes. The table covers team-class students with an end-of-year score. Collaborations and help received are per student-cycle. The last three columns are the two terms of (26) and their sum.

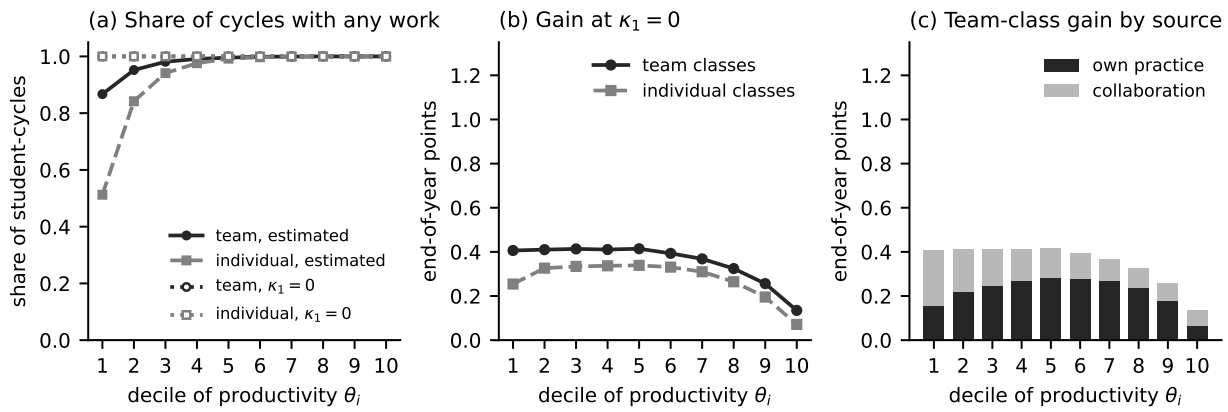


Figure C.2. A participation nudge by productivity

Notes. The figure groups the students of each arm by decile of productivity θ_i. Panel (a) shows the share of student-cycles with any work at the estimated κ_1 and at $\kappa_1 = 0$, panel (b) the gain in end-of-year points from removing the linear cost, $G^r(0)$, and panel (c) the team-class gain split into the parts due to own practice and to collaboration.

D Model details

This appendix derives the recognition probabilities (10) and (11) and the practice rule (22) of Section 5, using only the parameters of that section.

D.1 The recognition probability

This subsection derives the probability that a student is recognized under each recognition rule. As Appendix B.4 documents, a tournament records a score as the share of questions answered correctly, in points out of 100, so in a class-cycle whose tournament has Q_{ct} questions the recorded scores move in steps of $g_{ct} = 100/Q_{ct}$. Table B.8 reports these steps, which are read off each class-cycle's recorded scores. Because the scores in the model are continuous while the recorded ones move in steps, two students can land on the same recorded score and tie, and j finishes above i only when j's recorded score is a step higher. Write g for the class-cycle's step, fix i's own score at a value y_i, and let $D = Y_{jt} - y_i$ be how far classmate j's score is above it. Students evaluate the probability that j finishes strictly above i by averaging over a uniform position within a rounding interval,

$$\begin{aligned}\Pr(j \text{ strictly above } i) &= \frac{1}{g}\,\mathbb{E}\big[D\,\mathbf{1}\{0 < D < g\}\big] + \Pr(D > g) \\ &= \frac{m\left[\Phi\Big(\frac{g-m}{\sigma}\Big) - \Phi\Big(\frac{-m}{\sigma}\Big)\right] + \sigma\left[\phi\Big(\frac{-m}{\sigma}\Big) - \phi\Big(\frac{g-m}{\sigma}\Big)\right]}{g} + 1 - \Phi\Big(\frac{g-m}{\sigma}\Big),\end{aligned} \tag{D.1}$$

where Φ and ϕ are the standard normal distribution and density functions, and m and σ are the mean and the standard deviation of D, given below.

The first line has two terms, one for each way j can finish above i. To fix ideas, take a tournament with ten questions, so that $g = 10$. The recorded values are then 0, 10, 20 and so

on, and every underlying score is rounded to the nearest recorded value, so any underlying score from 45 up to 55 is recorded as 50, any score from 55 up to 65 as 60, and so on. The second term, $\Pr(D > g)$, covers the case where j's underlying score is more than ten points above i's. For example, if i is at 52 and j at 63, i is recorded as 50 and j as 60, and whatever the two underlying scores are, a gap of more than ten points always puts j on a higher recorded value. The first term covers the case where j is above i by less than ten points, $0 < D < g$. Now it depends on where the two scores fall. Suppose j is 4 points above i. If i is at 47 and j at 51, both are recorded as 50 and they tie. If i is at 53 and j at 57, i is recorded as 50 and j as 60, and j finishes above. So j finishes above exactly when i's underlying score lies within 4 points below a rounding point (55, 65, and so on), which happens 4 times in 10 if i's score is equally likely to fall anywhere between two rounding points; with a gap of D points the chance is D/g. Averaging D/g over the gaps between 0 and g gives the first term. If j is at or below i, j cannot finish above, which is why gaps $D \leq 0$ contribute nothing. The second line writes both terms out. Since j's score is normal by (7), D is normal with mean $m = \mathbb{E}[Y_{jt}] - y_i$ and standard deviation σ, and the first term is a partial expectation of a normal variable over $(0, g)$, which has a standard closed form. As $g \to 0$ the steps disappear, ties vanish, and the expression returns $\Phi(m/\sigma)$, the plain probability that j's score exceeds i's.

From the pairwise probability to the recognition probability. Under the individual recognition rule i is recognized when fewer than five attending classmates finish strictly above i. Hold i's own score at y_i. A classmate j counts against i when j attends and finishes above i, which happens with probability

$$q_{jt}(y_i) = \Lambda(\bar{u}^A_{jt}) \Pr(j \text{ strictly above } i \mid y_i), \tag{D.2}$$

the probability that j attends, which the attendance decision (15) determines jointly with the recognition probabilities, times the probability in (D.1). Treating these events as independent given y_i, the number of classmates who count is a sum of yes-or-no events with different probabilities, and its distribution is built one classmate at a time. Writing $P_n(c)$ for the probability that exactly c of the first n classmates count,

$$P_n(c) = P_{n-1}(c)\,\big(1 - q_{nt}\big) + P_{n-1}(c-1)\,q_{nt}, \qquad P_0(0) = 1, \tag{D.3}$$

since the n-th classmate either does not count, leaving the count at c, or counts, raising it from $c - 1$. For example, with two classmates who count with probabilities 0.5 and 0.2, nobody counts with probability $0.5 \times 0.8 = 0.40$, exactly one with $0.5 \times 0.8 + 0.5 \times 0.2 = 0.50$, and both with 0.10. With J classmates, i is recognized at score y_i with probability $\sum_{c=0}^{4} P_J(c)$. Finally, i's own score is random, so this is averaged over its distribution,

$$\widehat{\mathrm{EP}}^{\,\mathrm{ind}}_{it} = \int \sum_{c=0}^{4} P_J(c \mid y)\; f_{it}(y)\, dy, \tag{D.4}$$

where f_{it} is the normal density of Y_{it} from (7), with mean $\alpha_0 + \alpha_z z_i + k_{it}$ and standard deviation $\sqrt{\sigma_\alpha^2 + \sigma_Y^2}$, the same standard deviation σ that D has, and the integral is computed by Gauss–Hermite quadrature.

Under the team recognition rule the same construction applies to teams, with one more source of uncertainty, namely how many members of each team turn up. The total score of a team u is the sum of the scores of the members who attend, $T_{ut} = \sum_{j \in u} a_{jt} Y_{jt}$, and a member who attends adds their tournament score, including the intercept α_0, so a team with more members present has a higher total score. Let n_u be the number of members of u who attend; its distribution follows from the members' attendance probabilities by the recursion (D.3). Conditional on $n_u = n$, the total score is the sum of n scores, and students treat its distribution as normal, with the exact conditional mean and variance. The mean is the expected sum of the expected scores of the members present, given that n are present, and the variance is $n(\sigma_\alpha^2 + \sigma_Y^2)$ plus the variance of that sum of expected scores across the possible sets of n members present. A team with nobody present has no total score and cannot be recognized. Team u finishes strictly above a total score y of i's team with probability

$$p_{ut}(y) = \sum_{n \geq 1} \Pr(n_u = n)\ \Pr\big(T_{ut} > y \text{ strictly} \mid n_u = n\big), \tag{D.5}$$

where each conditional probability is (D.1) with m the conditional mean minus y and σ the conditional standard deviation. A team's total score is a sum of scores on a grid and lies on the same grid, so the same g applies. Student i's team is recognized when fewer than three other teams finish above it, so the recursion (D.3) runs over the other teams, with $p_{ut}(y)$ in place of q_{nt}, and the sum runs to $c = 2$. Finally the result is averaged over the total score of i's own team, in which i is present for certain and each teammate with their attendance probability. The number of teammates present, n', has its own distribution by (D.3), and given n' the total score of i's own team is treated in the same way, as a normal distribution whose mean is i's expected score plus the expected sum of the expected scores of the teammates present, and whose variance is $(1+n')(\sigma_\alpha^2+\sigma_Y^2)$ plus the variance of that sum across the possible sets of teammates present. The averaging uses Gauss–Hermite quadrature for each count, with the probability of finishing in the top three evaluated on a grid of values of the total score of i's own team. The derivative with respect to the knowledge of a teammate j, which the decisions of Section 5 use, is $\Lambda(\bar{u}_{jt}^A)$ times the derivative with respect to the total score of i's team computed with j present for certain, since j's knowledge enters that total only when j attends.

Two steps define how students evaluate these probabilities. The first averages the chance that a classmate finishes above i over where i's score falls between two recorded values, separately for each classmate, and then treats classmates as independent. The second treats a team's total score, given the number of members present, as normal. We compare the resulting probabilities with the recognition rule applied exactly, at the estimate, by simulating each class-cycle's tournament 3,000 times, drawing attendance and scores from the model, recording scores on the grid and applying the recognition rule exactly, ties included. The simulated frequency of recognition conditional on attending and the probability computed here differ by 0.008 per student-cycle on average, which equals the Monte Carlo error of the simulated frequency, and their means differ by 0.002 in each arm.

D.2 The first-order condition for practice

This subsection derives the practice rule (22) from the student's objective. Write $\tilde{c}_{it} = c_{it}(N_{it}/\bar{N})^{\psi}$ for the length-adjusted cost of time, so the objective in (19) is

$$U_{it}(e) = \beta^{e}\big(\theta_i\, e + \zeta\, \mathbb{E}[C_{it} \mid e]\big) + V_{it}^{A}(e) - \tilde{c}_{it}\Big(\kappa_1 e + \tfrac{e^{\xi}}{\xi}\Big). \tag{D.6}$$

Step 1. Knowledge. By (8), practice raises this cycle's knowledge directly, by θ_i per unit, and through the collaboration it attracts, by ζ per collaboration. Help received from teammate j is e_{it} times a logit whose index the student takes as given, by (17), so expected collaboration is linear in e and

$$\frac{dk_{it}}{de} = \theta_i + \zeta\, \frac{\partial \mathbb{E}[C_{it} \mid e]}{\partial e}, \qquad \frac{\partial \mathbb{E}[C_{it} \mid e]}{\partial e} = \sum_{j \neq i} \Lambda\big(\text{index of } j \text{ helping } i\big), \tag{D.7}$$

the second being the help i receives per unit of the assignment attempted, with the index taken from (17) at the cycle's attendance probabilities. Each collaboration also adds ζ to the knowledge of the teammate who helps, so the knowledge of teammate j rises by $dk_{jt}/de = \zeta\, \Lambda(\text{index of } j \text{ helping } i)$.

Step 2. The tournament. Practice reaches $V_{it}^{A} = \log(1+e^{\bar{u}_{it}^{A}})$ only through the attendance index $\bar{u}_{it}^{A}$ of (14); the cost of time in that index does not depend on e. The index depends on e through $B_v\, \widehat{\mathrm{EP}}_{it}$, and the recognition probability depends on e through i's knowledge and, under the team recognition rule, through the knowledge of the teammates who help i. The chain rule gives

$$\begin{aligned} \frac{dV_{it}^{A}}{de} &= \underbrace{\frac{\partial V_{it}^{A}}{\partial \bar{u}_{it}^{A}}}_{\Lambda(\bar{u}_{it}^{A})}\, \underbrace{\frac{\partial \bar{u}_{it}^{A}}{\partial \widehat{\mathrm{EP}}_{it}}}_{B_v} \Big(\frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial k_{it}}\, \frac{dk_{it}}{de} + \sum_{j \neq i} \frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial k_{jt}}\, \frac{dk_{jt}}{de}\Big) \\ &= \Lambda(\bar{u}_{it}^{A})\, B_v \Big[\frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial k_{it}}\Big(\theta_i + \zeta\, \frac{\partial \mathbb{E}[C_{it} \mid e]}{\partial e}\Big) + \zeta \sum_{j \neq i} \Lambda(\text{index of } j \text{ helping } i)\, \frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial k_{jt}}\Big], \end{aligned} \tag{D.8}$$

The first factor follows by differentiating $\log(1+e^{\bar{u}_{it}^{A}})$ with respect to $\bar{u}_{it}^{A}$. In words, a unit of practice adds dk/de to knowledge, each unit of knowledge raises the recognition chance by $\partial \widehat{\mathrm{EP}}_{it}/\partial k_{it}$, each unit of recognition chance raises the attendance index by B_v, and each unit of the index is worth $\Lambda(\bar{u}_{it}^{A})$, the probability of attending, because recognition is collected only by a student who turns up. The sum applies the same chain to the teammates who help i. Each collaboration adds ζ to the helper's knowledge as well, and under the team recognition rule that knowledge enters i's recognition probability through the team's total score when the helper attends, which $\partial \widehat{\mathrm{EP}}_{it}/\partial k_{jt}$ takes into account. In individual classes there is no helping, and the sum is zero.

Step 3. Marginal value and marginal cost. Differentiating (D.6) gives

$$U_{it}'(e) = mv_{it} - \tilde{c}_{it}\big(\kappa_1 + e^{\xi-1}\big), \tag{D.9}$$

where the marginal value of practice combines the knowledge benefit from Step 1 and the tournament benefit from Step 2,

$$\begin{aligned} mv_{it} &\equiv \beta^e \frac{dk_{it}}{de} + \frac{dV_{it}^A}{de} \\ &= \left(\theta_i + \zeta \, \frac{\partial \mathbb{E}[C_{it} \mid e]}{\partial e}\right)\left(\beta^e + \Lambda(\bar{u}_{it}^A)\, B_v \, \frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial k_{it}}\right) \\ &\quad + \Lambda(\bar{u}_{it}^A)\, B_v \, \zeta \sum_{j \neq i} \Lambda(\text{index of } j \text{ helping } i) \, \frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial k_{jt}}. \end{aligned}$$

The first factor is the additional knowledge produced by a unit of practice, θ_i from the student's own work plus $\zeta \, \partial \mathbb{E}[C_{it} \mid e]/\partial e$ from the help it attracts. The second factor is the value of each additional unit of knowledge, the direct value β^e plus the expected recognition benefit $\Lambda(\bar{u}_{it}^A) B_v \, \partial \widehat{\mathrm{EP}}_{it}/\partial k_{it}$. The last term is the recognition value of the knowledge that i's helpers gain, from Step 2. The subtracted term in (D.9), $\tilde{c}_{it}(\kappa_1 + e^{\xi-1})$, is the marginal cost of practice.

Step 4. Candidate solutions. At an interior solution, setting (D.9) equal to zero gives $e^{\xi-1} = mv_{it}/\tilde{c}_{it} - \kappa_1$. At $e = 0$, the derivative must be nonpositive, which requires $mv_{it}/\tilde{c}_{it} \leq \kappa_1$. At $e = 1$, the derivative must be nonnegative, which requires $mv_{it}/\tilde{c}_{it} \geq \kappa_1 + 1$. Marginal value is evaluated at the candidate share in each condition. Since $\xi > 1$, marginal cost rises with e. The decision rule evaluates the objective (D.6) at the interior candidate and at the two endpoints and selects the share with the highest value.

The linear part of the cost shapes the distribution of practice in two ways. First, a student whose marginal value per unit of cost falls short of κ_1 does nothing, and one just above it does a little, so practice falls continuously to zero and some student-cycles lie close to zero, 2.6% of them between 0 and 0.25. Second, it adds dispersion. Table B.5 shows that 32.8% of student-cycles are at full completion. Without the linear part, interior practice would be lognormal, and a lognormal distribution that matches this share and the interior mean of $\log e$, -0.522, has an interior variance of $\log e$ of 0.133, against 0.181 in the data. The linear part accounts for the difference.

E Model fit

This appendix defines the twenty-two moments of Table 8 and reports how well the model fits them. Table E.1 compares each targeted moment in the data and in the model, and Figure E.1 compares the distribution of practice in the data and in the model. The moments of tournament scores, practice and attendance are computed on the 3,481 student-cycles of the individual arm, except the score on practice, which uses all 7,372 student-cycles of the structural panel. The helping moments use the 3,891 student-cycles of the team arm, and the end-of-year moments use students with both scores. Simulated moments use the ninety-six draws and sixty quadrature nodes of the estimation, and simulated practice is recorded as in the data, as a whole number of questions attempted out of N_{it}.

The score moments use the tournament scores of attending students. The score on

practice is the coefficient on e_{it} in column (1) of Appendix Table B.3, a regression of the score on e_{it} and C_{it} with student and class-by-cycle fixed effects. The score on baseline and $\log e$ on baseline are least-squares slopes on the standardized baseline score z_i, and the between-student share of score variance is the variance of student means over the total variance. The moments of $\log e$ use cycles with $0 < e_{it} < 1$, and its within- and between-student variances are the variances of deviations from student means and of student means. Practice on assignment length is the within-student slope of e_{it} on $\log N_{it}$. Attendance on $\log e$ is the least-squares slope of attendance on $\log e_{it}$ over interior cycles, with both in deviations from the student's means over those cycles, and attendance on baseline is the slope of attendance on z_i within class and cycle.

Help given is the number of reported helping instances per team student-cycle. Its slope is the coefficient on the number of teammates shown behind, F^L_{jt}, in a pooled least-squares regression of help given on F^L_{jt}, the number of teammates shown ahead, F^H_{jt}, the student's own practice and the student's baseline score relative to teammates. A teammate is shown ahead when the teammate worked on the cycle's assignment and is not shown behind. Whether a teammate is shown behind is taken from the recorded reports, in the data and in the simulation alike, and whether a teammate worked is observed in the data and simulated in the model, like every other behavioral outcome.

The control end-of-year level is the mean end-of-year score of control students, and the control-arm slope and dispersion are the coefficient on S_i^{base} and the standard deviation of the residual in a regression of control students' end-of-year scores on S_i^{base} and on school and grade indicators. As Section 7 describes, the individual–control and team–control contrasts are the differences between each arm and control in mean residuals from a regression of end-of-year scores on the baseline score and on school and grade indicators, run on students in all three arms with both scores.

Untargeted moments provide a check of the model. Column (1) of Table B.3 reports the within-student coefficient of the tournament score on collaboration, which is 0.56 in the data and 0.48 in the model, 0.3 bootstrap standard errors apart. The coefficient differs from ζ because productivity varies across students and, within a student, practice and collaboration move together. The associations of practice and attendance with the mean baseline score of randomly assigned teammates are small. In the data, the association is -0.010 (0.028) for practice and 0.007 (0.035) for attendance, and in the model it is -0.017 for practice and -0.043 for attendance. Among attending students, the model recognizes a share 0.233 of students in individual classes, against 0.268 in the data, and 0.410 in team classes, against 0.420 in the data. Table E.2 reports standard errors from a bootstrap that resamples classes, the unit of randomization, and further untargeted moments, among them the differences in practice and attendance between the two platform arms, which the model fits less closely.

Table E.1. Model fit for the twenty-two targeted moments

Moment	Data	Model	(Model − Data)/SE
Score on practice, within student	1.848	1.655	−0.15
Team − control, end of year	3.353	3.350	−0.01
Individual − control, end of year	1.063	1.054	−0.01
Score level	36.587	36.586	0.00
Score on baseline	5.500	5.501	0.00
Between-student share of score variance	0.613	0.613	0.01
Score dispersion	20.656	20.661	0.01
Mean of $\log e$, interior	−0.517	−0.517	0.00
Share with $e = 0$	0.067	0.067	0.01
Share with $e = 1$	0.336	0.336	0.00
Within-student variance of $\log e$	0.068	0.068	−0.04
Between-student variance of $\log e$	0.123	0.123	0.03
$\log e$ on baseline	−0.005	−0.005	0.00
Practice on assignment length, within student	−0.089	−0.089	0.00
Attendance rate	0.688	0.688	−0.02
Attendance on $\log e$, within student, interior	0.094	0.095	0.02
Attendance on baseline	0.016	0.016	0.01
Help given, slope on teammates shown behind	0.328	0.327	−0.03
Help given, level	0.718	0.719	0.03
Control end-of-year level	41.016	41.028	0.02
Control end-of-year slope on baseline, net of school and grade	0.222	0.222	0.00
Control end-of-year dispersion, net of school and grade	13.498	13.510	0.03

Notes. Each moment of Table 8 is computed identically on the data and on the simulation at the estimate, with the ninety-six draws and sixty quadrature nodes of the estimation. The last column divides the difference by the bootstrap standard error of the data moment.

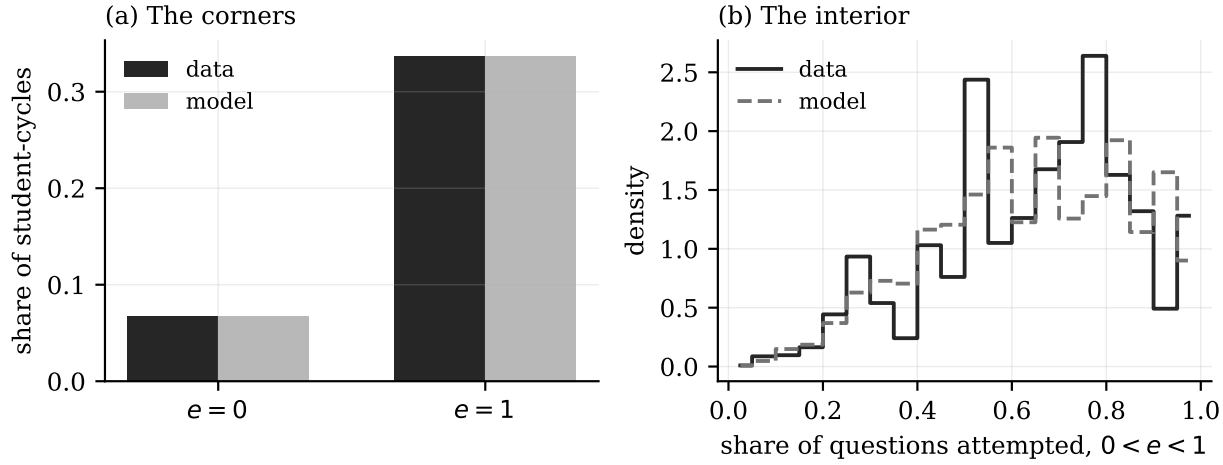


Figure E.1. The distribution of practice, data and model

Notes. The figure uses the student-cycles of the individual arm, the sample of the targeted practice moments. Model quantities are the average of the ninety-six simulation draws of the estimation at the estimate of Table 9. Panel (a) shows the shares of student-cycles with no work, $e = 0$, and with full completion, $e = 1$, and panel (b) shows the density of practice on the interior, $0 < e < 1$, in twenty bins.

Table E.2. Class-bootstrap standard errors and selected untargeted moments

	Estimate	Student-bootstrap SE	Class-bootstrap SE
Collaboration return, ζ	0.441	0.131	0.275
Value of helping a teammate shown behind, τ	0.646	0.284	0.442
Linear practice cost, κ_1	0.089	0.172	0.208
	Data	Model	Class-standardized gap
Team–individual difference in practice	−0.025	0.070	4.66
Team–individual difference in attendance	−0.015	0.050	1.98
Help given, slope on relative baseline	0.0043	0.0003	−3.30
Help given, slope on teammates shown ahead	0.079	0.232	3.48
Help received, slope on teammates shown behind	0.190	0.154	−0.69

Notes. The class bootstrap resamples classes within arms for 400 replications and holds the parameter estimate and the weighting matrix fixed. The top panel computes standard errors with the formula of Section 7, using the class-bootstrap covariance of the data moments. The bottom panel divides the difference between the model and the data by the class-bootstrap standard error. The team–individual differences in practice and attendance are raw differences in means on the structural sample, so the difference in practice differs from the regression-adjusted estimate in Table 7.

F Proofs of the identification propositions

This appendix proves the three identification propositions of Section 6. Each moment is defined, as a population quantity, in the step that uses it. A bar over a student-cycle variable denotes its mean over the cycles in which it is observed for that student. Write $\ell_{it} = \log(N_{it}/\bar{N})$, $x_{it} = mv_{it}/\{c_{it}(N_{it}/\bar{N})^{\psi}\}$, $r_{it} = \Lambda(\bar{u}^{A}_{it})B_v\,\partial\widehat{\mathrm{EP}}_{it}/\partial k_{it}$ for the recognition value of a unit of knowledge, and $h_{it} = \partial\mathbb{E}[C_{it} \mid e]/\partial e$ for the help attracted per unit of practice, which is zero in individual classes. By (20), $mv_{it} = (\theta_i + \zeta h_{it})(\beta^e + r_{it}) = \theta_i(1 + \zeta h_{it}/\theta_i)(\beta^e + r_{it})$, and by (12), $\log c_{it} = \pi_b z_i + \sigma_\nu \nu_{it}$. Taking logs of x_{it} and substituting $\log\theta_i = \theta_0 + \sigma_\theta u^{\theta}_i$ from (9),

$$\log x_{it} = \theta_0 + \sigma_\theta u^{\theta}_i + \log(\beta^e + r_{it}) + \log\Big(1 + \frac{\zeta h_{it}}{\theta_i}\Big) - \pi_b z_i - \psi\ell_{it} - \sigma_\nu\nu_{it}, \tag{F.1}$$

and by (22), $e_{it} = 0$ when $x_{it} \le \kappa_1$, $e_{it} = 1$ when $x_{it} \ge \kappa_1 + 1$, and $e^{\xi-1}_{it} = x_{it} - \kappa_1$ otherwise. Steps (B3) and (C3) below use one fact about least-squares slopes. The least-squares slope of a differentiable function of z on z equals a weighted average of the function's derivative, with nonnegative weights that integrate to one (Yitzhaki, 1996).

Proof of Proposition 1. The proof has three parts. Part A takes the practice and attendance decisions and σ_θ as given and inverts the score moments for θ_0, α_0, α_z, σ_α and σ_Y. Part B takes the score parameters and the recognition value r_{it} as given and inverts the practice moments for the six practice parameters and σ_θ. Part C takes the recognition probabilities as given and inverts the attendance moments for the three attendance parameters. All moments are computed on the individual arm, except the within-student slope of the score on practice, which uses both platform arms. In the individual arm $h_{it} = 0$, so (F.1) reduces to $\log x_{it} = \theta_0 + \sigma_\theta u^{\theta}_i + \log(\beta^e + r_{it}) - \pi_b z_i - \psi\ell_{it} - \sigma_\nu\nu_{it}$.

Part A. Tournament scores. On attending student-cycles, by (7) and (8),

$$Y_{it} = \alpha_0 + \alpha_z z_i + \theta_i e_{it} + \zeta C_{it} + \sigma_\alpha\omega_i + \sigma_Y\varepsilon^{Y}_{it}, \tag{F.2}$$

with $C_{it} = 0$ in the individual arm. The two shocks have mean zero and unit variance and are independent of z_i, of practice, of collaboration and of attendance, which are decided before the score is drawn (Assumption 2).

(A1) Productivity level θ_0. This step uses attending student-cycles in both platform arms, with C_{it} observed. Let e^{*}_{it} be the residual of practice on collaboration and on student and class-by-cycle fixed effects over these student-cycles. The difference $e_{it} - e^{*}_{it}$ lies in the span of the regressors, so $\mathbb{E}[e^{*}_{it}(e_{it} - e^{*}_{it})] = 0$ and $\mathbb{E}[e^{*}e] = \mathbb{E}[(e^{*})^2]$. The within-student slope of the score on practice, with collaboration as a covariate, is $b_e = \mathbb{E}[e^{*}Y]/\mathbb{E}[(e^{*})^2]$. Multiplying (F.2) by e^{*}_{it} and taking expectations term by term,

$$\mathbb{E}[e^{*}Y] = \alpha_0\mathbb{E}[e^{*}] + \alpha_z\mathbb{E}[e^{*}z_i] + \mathbb{E}[e^{*}\theta_i e_{it}] + \zeta\mathbb{E}[e^{*}C_{it}] + \sigma_\alpha\mathbb{E}[e^{*}\omega_i] + \sigma_Y\mathbb{E}[e^{*}\varepsilon^{Y}_{it}]. \tag{F.3}$$

The first, second and fifth terms are zero because e^{*} sums to zero within each student while 1, z_i and ω_i are constant within student. The fourth term is zero because C_{it} is a regressor in the residualization. The last term is zero because ε^{Y}_{it} has mean zero and is independent

of practice and collaboration. Hence

$$b_e = \frac{\mathbb{E}[e^*_{it} e_{it}\, \theta_i]}{\mathbb{E}[(e^*_{it})^2]}, \tag{F.4}$$

a weighted average of productivity across students in which student i carries the weight $\mathbb{E}[e^*_{it} e_{it} \mid i]/\mathbb{E}[(e^*_{it})^2]$. These weights average to one because $\mathbb{E}[e^* e] = \mathbb{E}[(e^*)^2]$, and a weight can be negative for a student whose practice moves against the class-by-cycle effects. Substituting $\theta_i = e^{\theta_0 + \sigma_\theta u_i^\theta}$ gives $b_e = e^{\theta_0}\, \mathbb{E}[e^*_{it} e_{it} e^{\sigma_\theta u_i^\theta}]/\mathbb{E}[(e^*_{it})^2]$. The expectation involves the joint distribution of practice and productivity, which the model generates. Part A takes the practice decisions as given, so this distribution, and with it the expectation, does not vary with θ_0. For team students this distribution also reflects the help that practice attracts, and the step takes their practice decisions as given, as it does for individual-arm students. When $b_e > 0$, the expectation is positive at any parameter vector at which the model matches b_e, and given σ_θ the right-hand side is then strictly increasing in θ_0, so

$$\theta_0 = \log b_e - \log \frac{\mathbb{E}[e^*_{it} e_{it} e^{\sigma_\theta u_i^\theta}]}{\mathbb{E}[(e^*_{it})^2]}. \tag{F.5}$$

(A2) Level α_0 and baseline gradient α_z. Write $k_{it} = \theta_i e_{it}$ for the knowledge of an individual-arm student. Taking the mean of (F.2) over attending student-cycles of the individual arm, the two shocks drop out because they have mean zero, so $\mathbb{E}[Y_{it}] = \alpha_0 + \alpha_z \mathbb{E}[z_i] + \mathbb{E}[k_{it}]$. Taking the covariance with z_i, the two shocks drop out because they are independent of z_i, so $\mathrm{Cov}(Y_{it}, z_i) = \alpha_z \mathrm{Var}(z_i) + \mathrm{Cov}(k_{it}, z_i)$. Solving the second equation for α_z and then the first for α_0,

$$\alpha_z = \frac{\mathrm{Cov}(Y_{it}, z_i) - \mathrm{Cov}(k_{it}, z_i)}{\mathrm{Var}(z_i)}, \qquad \alpha_0 = \mathbb{E}[Y_{it}] - \alpha_z \mathbb{E}[z_i] - \mathbb{E}[k_{it}], \tag{F.6}$$

where the moments of k_{it} are computed under the model from the productivity distribution and the practice decisions.

(A3) Permanent and transitory dispersion σ_α and σ_Y. Consider the variance of Y_{it} across attending student-cycles of the individual arm. In (F.2) the term $\alpha_z z_i + k_{it}$, the shock $\sigma_\alpha \omega_i$ and the shock $\sigma_Y \varepsilon^Y_{it}$ are mutually uncorrelated, by the independence in Assumption 2, so the variance of the sum is the sum of the variances,

$$\mathrm{Var}(Y_{it}) = \mathrm{Var}(\alpha_z z_i + k_{it}) + \sigma^2_\alpha + \sigma^2_Y. \tag{F.7}$$

Let $\bar{Y}_i$ be student i's mean score over the T_i tournaments attended. The between-student share of score variance is $\mathrm{Var}(\bar{Y}_i)/\mathrm{Var}(Y_{it})$, both variances taken across student-cycles. Averaging (F.2) over the student's attended cycles gives $\bar{Y}_i = \alpha_0 + \alpha_z z_i + \bar{k}_i + \sigma_\alpha \omega_i + \sigma_Y \bar{\varepsilon}^Y_i$, where $\bar{\varepsilon}^Y_i$ is the mean of T_i independent standard normal draws, so $\mathrm{Var}(\bar{\varepsilon}^Y_i \mid T_i) = 1/T_i$. A variance across student-cycles counts each student T_i times, so the transitory shock contributes the average over student-cycles of σ^2_Y/T_i, in which each student's term σ^2_Y/T_i appears T_i times. The T_i cancel, and the contribution is σ^2_Y times the number of students divided by

the number of student-cycles,

$$\frac{\sum_i T_i \left(\sigma_Y^2/T_i\right)}{\sum_i T_i} = \frac{\sigma_Y^2}{\bar{T}}, \tag{F.8}$$

where $\bar{T}$ is the mean number of attended tournaments per student. The permanent shock contributes σ_α^2 in full, and the three components of $\bar{Y}_i$ are again uncorrelated, so

$$\mathrm{Var}(\bar{Y}_i) = \mathrm{Var}(\alpha_z z_i + \bar{k}_i) + \sigma_\alpha^2 + \frac{\sigma_Y^2}{\bar{T}}. \tag{F.9}$$

Equations (F.7) and (F.9) are two linear equations in σ_α^2 and σ_Y^2. Subtracting the second from the first eliminates σ_α^2 and leaves $\sigma_Y^2(1 - 1/\bar{T})$ on the right, so

$$\sigma_Y^2 = \frac{\{\mathrm{Var}(Y_{it}) - \mathrm{Var}(\alpha_z z_i + k_{it})\} - \{\mathrm{Var}(\bar{Y}_i) - \mathrm{Var}(\alpha_z z_i + \bar{k}_i)\}}{1 - 1/\bar{T}}, \qquad \sigma_\alpha^2 = \mathrm{Var}(Y_{it}) - \mathrm{Var}(\alpha_z z_i + k_{it}) - \sigma_Y^2, \tag{F.10}$$

where the two knowledge variances are computed under the model. The division requires $\bar{T} > 1$, that is, students attend more than one tournament on average.

Part B. Practice. Let μ_{it} denote the right-hand side of (F.1) without the cost shock, so that $\log x_{it} = \mu_{it} - \sigma_\nu \nu_{it}$. Part B takes the equilibrium recognition value r_{it} as given. Because r_{it} depends on the attendance probability $\Lambda(\bar{u}_{it}^A)$, which depends on the cost of time, and on the student's own knowledge, r_{it} and with it μ_{it} can vary with the cycle's cost shock ν_{it}. Each step therefore proceeds in two stages. It first derives the relation between its moments and its parameters in closed form when r_{it} does not vary with ν_{it}, in which case, given θ_i, z_i, ℓ_{it} and r_{it}, the cost shock is the only random term in $\log x_{it}$ and is standard normal (Assumption 2). It then treats the general case, in which each moment is a known function of the parameters given the equilibrium objects, computed under the model, and the moments determine the parameters when their Jacobian with respect to these parameters is nonsingular.

(B1) Linear cost κ_1 and value of knowledge β^e. Practice is recorded as a whole number of questions out of N_{it}, so recorded practice is zero when $e_{it} < 1/(2N_{it})$ and one when $e_{it} \geq 1 - 1/(2N_{it})$. By the decision rule, the first event is $x_{it} - \kappa_1 < (2N_{it})^{1-\xi}$ and the second is $x_{it} - \kappa_1 \geq (1 - 1/2N_{it})^{\xi - 1}$. Write $\underline{\kappa}_{it} = \kappa_1 + (2N_{it})^{1-\xi}$ and $\bar{\kappa}_{it} = \kappa_1 + (1 - 1/2N_{it})^{\xi-1}$ for the two thresholds, so recorded practice is zero when $x_{it} < \underline{\kappa}_{it}$ and one when $x_{it} \geq \bar{\kappa}_{it}$. Holding the equilibrium objects fixed, x_{it} rises with β^e at every value of the cost shock and both thresholds rise with κ_1. The share with no work is therefore increasing in κ_1 and decreasing in β^e, and the share with full completion moves in the opposite directions, strictly so because ν_{it} is normal and x_{it} varies continuously with it. When r_{it} does not vary with ν_{it}, recorded practice is zero when $\mu_{it} - \sigma_\nu \nu_{it} < \log \underline{\kappa}_{it}$, that is, when $\nu_{it} > (\mu_{it} - \log \underline{\kappa}_{it})/\sigma_\nu$, an event of probability $\Phi\big((\log \underline{\kappa}_{it} - \mu_{it})/\sigma_\nu\big)$ since ν_{it} is standard normal, where Φ is the standard normal distribution function. Likewise recorded practice is one when $\nu_{it} \leq (\mu_{it} - \log \bar{\kappa}_{it})/\sigma_\nu$. Hence

$$\Pr(e_{it} = 0 \mid \cdot) = \Phi\Big(\frac{\log \underline{\kappa}_{it} - \mu_{it}}{\sigma_\nu}\Big), \qquad \Pr(e_{it} = 1 \mid \cdot) = \Phi\Big(\frac{\mu_{it} - \log \bar{\kappa}_{it}}{\sigma_\nu}\Big). \tag{F.11}$$

The share with no work is the average of the first probability over student-cycles, and the share with full completion is the average of the second. Every assignment has more than

one question, so $(2N_{it})^{1-\xi} \leq 4^{1-\xi} < (3/4)^{\xi-1} \leq (1-1/2N_{iu})^{\xi-1}$, and every threshold $\underline{\kappa}_{it}$ lies below every threshold $\bar{\kappa}_{iu}$. When r_{it} takes a common value r, the two shares determine (κ_1, β^e) given ξ and the remaining parameters. Suppose that $(\kappa_1', \beta^{e\prime})$ with $\kappa_1' > \kappa_1$ gave the same two shares. Since the share with no work rises with κ_1 and falls with β^e, $\beta^{e\prime} > \beta^e$, and μ_{it} rises by $D = \log\{(\beta^{e\prime}+r)/(\beta^e+r)\} > 0$ in every student-cycle. The argument of the first probability in (F.11) changes by $\{\log(\underline{\kappa}_{it}'/\underline{\kappa}_{it}) - D\}/\sigma_\nu$, so an unchanged share with no work requires $\log(\underline{\kappa}_{it}'/\underline{\kappa}_{it}) \leq D$ in some student-cycle. The argument of the second changes by $\{D - \log(\bar{\kappa}_{iu}'/\bar{\kappa}_{iu})\}/\sigma_\nu$, so an unchanged share with full completion requires $D \leq \log(\bar{\kappa}_{iu}'/\bar{\kappa}_{iu})$ in some student-cycle. Both thresholds rise by $\kappa_1' - \kappa_1$, so $\log(\underline{\kappa}_{it}'/\underline{\kappa}_{it}) = \log\{1 + (\kappa_1' - \kappa_1)/\underline{\kappa}_{it}\}$, which exceeds $\log\{1 + (\kappa_1' - \kappa_1)/\bar{\kappa}_{iu}\} = \log(\bar{\kappa}_{iu}'/\bar{\kappa}_{iu})$ for every pair of student-cycles, a contradiction. The case $\kappa_1' < \kappa_1$ reverses every inequality, and $\kappa_1' = \kappa_1$ with $\beta^{e\prime} \neq \beta^e$ changes the share with no work. When r_{it} varies across student-cycles or with ν_{it}, β^e shifts $\log x_{it}$ by different amounts, and the two shares determine (κ_1, β^e) when their Jacobian with respect to these two parameters is nonsingular.

(B2) Curvature ξ. Steps (B2) to (B4) and (C2) use the share attempted on interior cycles. The data record a whole number of questions, which locates the share attempted within $1/(2N_{it})$ of the recorded value, and the estimator records simulated practice in the same way, so each matched moment is computed identically on the data and on the simulation. On the interior, $e_{it}^{\xi-1} = x_{it} - \kappa_1$, so $\log e_{it} = \log(x_{it} - \kappa_1)/(\xi - 1)$ with $x_{it} - \kappa_1 \in (0, 1)$. Given the equilibrium objects, the interior set and the distribution of x_{it} on it depend on κ_1, β^e, π_b, ψ, σ_ν and r_{it}, and ξ enters only the map from x_{it} to e_{it}, whether or not r_{it} varies with ν_{it}. Taking the mean over interior student-cycles, $\mathbb{E}[\log e_{it} \mid 0 < e_{it} < 1] = \mathbb{E}[\log(x_{it} - \kappa_1) \mid 0 < e_{it} < 1]/(\xi - 1)$, and solving for ξ,

$$\xi = 1 + \frac{\mathbb{E}[\log(x_{it} - \kappa_1) \mid 0 < e_{it} < 1]}{\mathbb{E}[\log e_{it} \mid 0 < e_{it} < 1]}, \tag{F.12}$$

where the numerator is computed under the model and the denominator is the mean of interior log practice. Both are negative, so $\xi > 1$. The two corner shares determine (κ_1, β^e) given ξ, and the interior mean determines ξ given (κ_1, β^e), so the three moments determine the three parameters together when their Jacobian with respect to the three parameters is nonsingular.

(B3) Cost shifters π_b and ψ. Holding r_{it} fixed, on the interior $\log e_{it} = \log(x_{it} - \kappa_1)/(\xi - 1)$, and by (F.1), $\partial \log x_{it}/\partial z_i = -\pi_b$ and $\partial \log x_{it}/\partial \ell_{it} = -\psi$. By the chain rule, $\partial \log e_{it}/\partial z_i = \{1/(\xi-1)\}\{1/(x_{it} - \kappa_1)\}\, \partial x_{it}/\partial z_i$ with $\partial x_{it}/\partial z_i = x_{it}\, \partial \log x_{it}/\partial z_i = -\pi_b x_{it}$, and $\partial e_{it}/\partial \ell_{it} = e_{it}\, \partial \log e_{it}/\partial \ell_{it}$, so

$$\frac{\partial \log e_{it}}{\partial z_i} = -\pi_b \frac{x_{it}}{(\xi - 1)(x_{it} - \kappa_1)}, \qquad \frac{\partial e_{it}}{\partial \ell_{it}} = -\psi \frac{e_{it}\, x_{it}}{(\xi - 1)(x_{it} - \kappa_1)}, \tag{F.13}$$

and at the corners both derivatives are zero. The moments are least-squares slopes on the interior sample, which is selected on x_{it}. When r_{it} does not vary with ν_{it}, on the interior $\log x_{it}$ lies between $\log \kappa_1$ and $\log(\kappa_1 + 1)$ with density proportional to $\exp\{-(\log x_{it} - \mu_{it})^2/2\sigma_\nu^2\}$, which depends on z_i only through μ_{it}, and a lower μ_{it} shifts this density toward lower values of $\log x_{it}$ in the monotone likelihood ratio order. Given ℓ_{it} and r_{it}, the conditional mean of interior log practice is therefore strictly decreasing in $\pi_b z_i$. When assignment length and r_{it}

do not vary with z_i, the least-squares slope of interior log practice on z_i is a positive-weighted average of the derivative of this conditional mean (Yitzhaki, 1996) and has the sign of $-\pi_b$. In general the slope also reflects how assignment length and r_{it} vary with the baseline score and how r_{it} varies with the cost shock, and it is a known function of π_b given ψ and the other parameters, computed under the model. The within-student slope of practice on ℓ_{it} uses the variation of assignment length across a student's cycles, which is independent of productivity and the cost shocks (Assumption 1), and it is likewise a known function of ψ. The two slopes determine π_b and ψ when their Jacobian with respect to these two parameters is nonsingular.

(B4) Productivity and cost dispersions σ_θ and σ_ν. Let $s^2 = \sigma_\theta^2 + \sigma_\nu^2$ and $d_{it} = \log(\beta^e + r_{it}) - \psi\ell_{it}$, which is known given β^e and ψ from steps (B1) and (B3) and r_{it}. By (F.1), $\log x_{it} = \theta_0 - \pi_b z_i + d_{it} + w_{it}$ with $w_{it} = \sigma_\theta u_i^\theta - \sigma_\nu \nu_{it}$. The composite shock w_{it} is normal with mean zero and variance s^2, independent of z_i and ℓ_{it} (Assumptions 1 and 2), and for two cycles t and u of the same student (w_{it}, w_{iu}) is bivariate normal with correlation σ_θ^2/s^2, because the cycles share u_i^θ and have independent cost shocks. On an interior cycle, $e_{it}^{\xi-1} = x_{it} - \kappa_1$, so $\log x_{it} = \log(\kappa_1 + e_{it}^{\xi-1})$ is observed given κ_1 and ξ, and hence so is

$$w_{it} = \log(\kappa_1 + e_{it}^{\xi-1}) - \theta_0 + \pi_b z_i - d_{it}, \tag{F.14}$$

given θ_0 from Part A and π_b from step (B3). A cycle is interior exactly when w_{it} lies in the interval $(\log\kappa_1 - \theta_0 + \pi_b z_i - d_{it},\ \log(\kappa_1 + 1) - \theta_0 + \pi_b z_i - d_{it})$. When r_{it} does not vary with ν_{it}, d_{it} does not depend on w_{it}, the endpoints are known, and the distribution of interior practice therefore reveals the distribution of w_{it} truncated to a known interval, and the joint distribution of interior practice in two cycles of the same student reveals the distribution of (w_{it}, w_{iu}) truncated to a known rectangle. The density of a normal variable truncated to a known interval is proportional to $\exp(-w^2/2s^2)$ on that interval, which determines s^2, and the density of a bivariate normal pair truncated to a known rectangle is proportional to $\exp\{-(w_{it}^2 - 2\rho w_{it}w_{iu} + w_{iu}^2)/2s^2(1-\rho^2)\}$, in which the coefficient on $w_{it}w_{iu}$ determines $\rho = \sigma_\theta^2/s^2$ given s^2. Hence $\sigma_\theta^2 = \rho s^2$ and $\sigma_\nu^2 = (1-\rho)s^2$. The argument uses students with two or more cycles with distinct interior practice (Assumption 4). The estimator matches two summaries of this distribution, the between- and within-student variances of interior log practice, that is, the variance of its student means and the variance of its deviations from them. In general, including when r_{it} varies with ν_{it}, both are known functions of σ_θ and σ_ν given the other parameters, computed under the model, and they determine σ_θ and σ_ν when their Jacobian with respect to these two parameters is nonsingular.

Part C. Attendance. Substituting $\log c_{it} = \pi_b z_i + \sigma_\nu \nu_{it}$ from (12) into (14), the attendance index is

$$\bar{u}_{it}^A = \mu_a - \rho_c \pi_b z_i - \rho_c \sigma_\nu \nu_{it} + B_v\big(\widehat{\mathrm{EP}}_{it} - \overline{\mathrm{EP}}\big), \tag{F.15}$$

and $\Pr(a_{it} = 1) = \Lambda(\bar{u}_{it}^A)$, with the recognition probability held at its equilibrium value.

(C1) Intercept μ_a. The attendance rate is $\mathbb{E}[\Lambda(\bar{u}_{it}^A)]$. Since Λ is strictly increasing and μ_a enters the index additively, the rate is strictly increasing in μ_a and determines it given the other terms.

(C2) Cost loading ρ_c. On an interior cycle, $\log x_{it}$ is observed as in step (B4), and by (F.1) the cost shock is $\sigma_\nu \nu_{it} = \theta_0 - \pi_b z_i + \sigma_\theta u_i^\theta + d_{it} - \log x_{it}$. Substituting this into (F.15), the

terms in z_i cancel and the attendance index on interior cycles is

$$\bar{u}^A_{it} = \big[\mu_a - \rho_c(\theta_0 + \sigma_\theta u^\theta_i)\big] + \rho_c\big(\log x_{it} - d_{it}\big) + B_v\big(\widehat{\mathrm{EP}}_{it} - \overline{\mathrm{EP}}\big). \tag{F.16}$$

The first term is constant within a student, $\log x_{it} - d_{it}$ is observed given the parameters of Part B, and the recognition probability is held at its equilibrium value. Attendance on interior cycles is therefore a logit with a student fixed effect in which the coefficient on the observed quantity $\log x_{it} - d_{it}$ is ρ_c. Such a logit identifies its slope coefficient from students whose regressor varies across cycles (Chamberlain, 1980), and a positive mass of students have two cycles with distinct interior practice (Assumption 4). The estimator matches the within-student least-squares slope of attendance on interior log practice, a linear summary of this relation in which $\log e_{it}$ is a strictly increasing function of $\log x_{it}$. The slope is a known function of ρ_c given the other parameters, computed under the model, and it determines ρ_c when its derivative with respect to ρ_c is nonzero.

(C3) Value of recognition B_v. Within class and cycle, the baseline score enters (F.15) through $-\rho_c\pi_b z_i$ and through $\widehat{\mathrm{EP}}_{it}$, which depends on z_i through the expected score $\alpha_0 + \alpha_z z_i + k_{it}$. Differentiating $\Lambda(\bar{u}^A_{it})$,

$$\frac{\partial \Pr(a_{it}=1)}{\partial z_i} = \Lambda'(\bar{u}^A_{it})\Big[-\rho_c\pi_b + B_v\frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial z_i}\Big]. \tag{F.17}$$

By the weighted-average property, the slope of attendance on the baseline score within class and cycle is $\mathbb{E}[\varpi_{it}\,\partial \Pr(a_{it}=1)/\partial z_i]$ for nonnegative weights ϖ_{it} that average to one. Substituting (F.17) and collecting terms,

$$\text{slope} = -\rho_c\pi_b\,\mathbb{E}\big[\varpi_{it}\Lambda'(\bar{u}^A_{it})\big] + B_v\,\mathbb{E}\Big[\varpi_{it}\Lambda'(\bar{u}^A_{it})\frac{\partial \widehat{\mathrm{EP}}_{it}}{\partial z_i}\Big], \tag{F.18}$$

where $\bar{u}^A_{it}$, and with it $\Lambda'(\bar{u}^A_{it})$ and the weights, also depend on B_v. With ρ_c from step (C2) and π_b from step (B3), the slope is a known function of B_v given the other parameters, computed under the model, in which B_v enters only through the recognition channel. This is the exclusion restriction behind B_v. The baseline score shifts attendance only through the cost of time, whose coefficient is known at this step, and through the chance of recognition. The slope determines B_v when its derivative with respect to B_v is nonzero.

Parts A, B and C each determine their own parameters from their own moments given the objects the other parts determine, namely the practice and attendance decisions and σ_θ in Part A, the score parameters and r_{it} in Part B, and the recognition probabilities in Part C. The fifteen parameters are the fixed point of the three parts. □

Proof of Proposition 2. For control students the practice and collaboration sums in (24) are zero and the control indicator is one, so the equation reduces to

$$S^{\text{end}}_i = \delta_0 + \delta_C + \lambda S^{\text{base}}_i + \gamma_{s(i)} + \gamma_{g(i)} + \varepsilon_i, \tag{F.19}$$

where $\gamma_{s(i)} + \gamma_{g(i)}$ is a linear combination of school and grade indicators that averages to zero among control students, and ε_i has mean zero and variance σ^2_ε and is independent of the

baseline score, school and grade (Section 5.6 and Assumption 2). Consider the least-squares regression, over control students, of S_i^{end} on a constant, S_i^{base} and school and grade indicators. In (F.19), $\delta_0 + \delta_C + \lambda S_i^{\text{base}} + \gamma_{s(i)} + \gamma_{g(i)}$ is a linear combination of these regressors and ε_i is uncorrelated with each of them, so the regression recovers that linear combination. Its coefficient on S_i^{base} is therefore λ, its residual is ε_i, and the standard deviation of the residual is σ_ε. Taking the mean of (F.19) over control students, $\gamma_{s(i)} + \gamma_{g(i)}$ averages to zero and ε_i has mean zero, so

$$\delta_0 + \delta_C = \mathbb{E}[S_i^{\text{end}} \mid \text{Control}_i = 1] - \lambda\, \mathbb{E}[S_i^{\text{base}} \mid \text{Control}_i = 1]. \tag{F.20}$$

The three control-arm moments are the coefficient on S_i^{base}, the standard deviation of the residual and the mean of S_i^{end}, so they give λ, then σ_ε, then the sum $\delta_0 + \delta_C$, which the proof of Proposition 3 separates into its two terms. □

Proof of Proposition 3. (1) Control-arm level δ_C and collaboration return ζ. Stack the students of the three arms who have both scores, let X be the matrix whose rows are a constant, S_i^{base} and the school and grade indicators, and let $M = I - X(X'X)^{-1}X'$, so that MV is the vector of residuals from the regression of V on X. Let a_I have entries $1/n_I$ for the n_I individual-arm students, $-1/n_C$ for the n_C control students and zero otherwise, and define a_T in the same way for the n_T team students and control students. The individual–control contrast is $a_I' M S^{\text{end}}$, the mean residual among individual-arm students minus the mean residual among control students, and the team–control contrast is $a_T' M S^{\text{end}}$. For any student-level variable V write $L_I(V) = a_I' MV$ and $L_T(V) = a_T' MV$. Both are linear in V, and because $MX = 0$ both are zero for a constant, for the baseline score and for any linear combination of school and grade indicators. Write $K_i = \sum_t \theta_i e_{it}$ for the knowledge a student builds from own practice and $A_i = \sum_t C_{it}$ for the number of collaborations, both zero for control students and A_i also zero for individual-arm students. Applying L_I and L_T to each term of (24),

$$L_r(S^{\text{end}}) = \delta_0 L_r(1) + \lambda L_r(S^{\text{base}}) + L_r(\gamma_s + \gamma_g) + \delta_C L_r(\text{Control}) + L_r(K) + \zeta L_r(A) + L_r(\varepsilon), \qquad r \in \{I, T\}. \tag{F.21}$$

Here $\gamma_s + \gamma_g$ is the vector with entries $\gamma_{s(i)} + \gamma_{g(i)}$. The first three terms are zero, the third because this vector is a linear combination of the school and grade indicators. The last has mean zero because ε_i is i.i.d. across students and independent of the arm, the baseline score, the school and the grade (Section 5.6 and Assumption 2). The collaboration term does not vanish for the individual arm, although $A_i = 0$ for every individual-arm and control student, because M is built from all three arms and the fitted values of A are nonzero for them. Writing $c_r \equiv L_r(\text{Control})$, the two contrasts are two linear equations in δ_C and ζ,

$$\begin{pmatrix} L_I(S^{\text{end}}) - L_I(K) \\ L_T(S^{\text{end}}) - L_T(K) \end{pmatrix} = \begin{pmatrix} c_I & L_I(A) \\ c_T & L_T(A) \end{pmatrix} \begin{pmatrix} \delta_C \\ \zeta \end{pmatrix}, \tag{F.22}$$

in which the contrasts of the end-of-year scores are observed, c_I, c_T, $L_I(A)$ and $L_T(A)$ are computed from the data, and $L_I(K)$ and $L_T(K)$ are computed under the model from the productivity distribution of Proposition 1 and the practice decisions, with team practice at its equilibrium value. The system determines δ_C and ζ when its determinant $c_I L_T(A) - c_T L_I(A)$

is nonzero. Then δ_0 follows from (F.20).

(2) Helping taste τ and cost κ_h. By (17), the probability that j helps teammate k in cycle t is $e_{kt}\Lambda(\tau I_{kt} + \chi_{jkt} - \kappa_h)$, where $\chi_{jkt} = \zeta\big[\Lambda(\bar{u}^A_{jt})B_v\big(\partial\widehat{\mathrm{EP}}_{jt}/\partial k_{jt} + \partial\widehat{\mathrm{EP}}_{jt}/\partial k_{kt}\big) + \beta^e\big]$ depends on ζ and on the parameters of Proposition 1. Help given is $O_{jt} = \sum_{k\neq j} h_{jkt}$, the number of teammates shown behind is $F^L_{jt} = \sum_{k\neq j} I_{kt}$, and the number shown ahead, F^H_{jt}, counts the teammates who worked and are not shown behind. Taking the expectation of O_{jt} over the helping shocks, the level of help given is

$$\mathbb{E}[O_{jt}] = \mathbb{E}\Big[\sum_{k\neq j} e_{kt}\Lambda(\tau I_{kt} + \chi_{jkt} - \kappa_h)\Big], \tag{F.23}$$

strictly decreasing in κ_h and strictly increasing in τ, since Λ is strictly increasing. The second moment is the coefficient on F^L_{jt} in a least-squares regression, over team student-cycles, of O_{jt} on a constant, F^L_{jt}, F^H_{jt}, the student's own practice and the student's baseline score relative to teammates. Let F^{L*}_{jt} be the residual of F^L_{jt} from the regression on the other regressors. By the identity used in step (A1) of the proof of Proposition 1, and replacing O_{jt} by its expectation over the helping shocks, which are independent of the regressors, the coefficient is

$$b_L = \frac{\mathbb{E}\big[F^{L*}_{jt}\sum_{k\neq j} e_{kt}\Lambda(\tau I_{kt} + \chi_{jkt} - \kappa_h)\big]}{\mathbb{E}\big[(F^{L*}_{jt})^2\big]}. \tag{F.24}$$

When $\chi_{jkt} = \chi$ and $e_{kt} = e$ are common across pairs, write $p_1 = \Lambda(\tau+\chi-\kappa_h)$ and $p_0 = \Lambda(\chi-\kappa_h)$ for the probabilities of helping a teammate shown behind and one shown ahead, and $\bar{F}^L$ and $\bar{F}^H$ for the mean numbers of each. The sum in (F.23) and (F.24) is then $e(p_1F^L_{jt}+p_0F^H_{jt})$. Since F^{L*}_{jt} is uncorrelated with F^H_{jt} and $\mathbb{E}[F^{L*}_{jt}F^L_{jt}] = \mathbb{E}[(F^{L*}_{jt})^2]$, the coefficient is $b_L = ep_1$ and the level is $e(p_1\bar{F}^L+p_0\bar{F}^H)$. Hence $p_1 = b_L/e$, $p_0 = \{\text{level} - b_L\bar{F}^L\}/(e\bar{F}^H)$, and inverting the logistic function,

$$\tau = \mathrm{logit}(p_1) - \mathrm{logit}(p_0), \qquad \kappa_h = \chi - \mathrm{logit}(p_0). \tag{F.25}$$

When χ_{jkt} and e_{kt} differ across pairs, the level (F.23) remains strictly decreasing in κ_h and strictly increasing in τ, and (F.24) is a known function of τ and κ_h given ζ and the parameters of Proposition 1, computed under the model. The two moments then determine τ and κ_h when their Jacobian with respect to these two parameters is nonsingular.

(3) Joint determination. The proposition takes the practice of team students at its equilibrium value, so $L_T(K)$ in step (1) does not vary with τ and κ_h. Step (1) therefore determines δ_C and ζ, and step (2) then determines τ and κ_h given ζ, through χ_{jkt}. □

Three conditions in the proofs concern the population moments, and each holds in the data. The within-student slope of the score on practice is positive, $b_e = 1.848$, as step (A1) requires. Students attend more than one tournament on average, as step (A3) requires. The determinant of the system (F.22) is -4.59, with $c_I = -0.946$, $c_T = -0.929$, $L_I(A) = -0.105$ and $L_T(A) = 4.753$, as step (1) of the proof of Proposition 3 requires.