Title: Primary- and middle-school teachers in South Asia overestimate the performance of their students

Authors: Sharnic Djaker[1]*, Alejandro J. Ganimian[1], Shwetlena Sabarwal[2]

Affiliations:

[1]New York University, Steinhardt School of Culture, Education, and Human Development; New York, United States.

[2]World Bank; Washington, D.C., United States.

*Corresponding author. Email: sharnic.djaker@nyu.edu

Abstract: Researchers have argued that teachers in developing countries do not devote enough attention to low-achieving students primarily because they face incentives to focus on their high-achieving peers. We use math and language achievement data for 1,500 students and survey data for 450 teachers across India and Bangladesh to highlight another potential explanation: most teachers do not know their students' academic skills. We show that many teachers underestimate the share of low performers in their classrooms, and that they believe that those students will perform better than they actually do. These results are not driven by less educated, trained, or experienced teachers or explained by biases against female, low-income, or lower caste students. Instead, teachers seem to overweight the importance of students' fluid intelligence.

One-Sentence Summary: Teachers in India and Bangladesh underestimate the extent to which their own students underperform in math and language.

There is mounting evidence that teachers in low- and middle-income countries rarely cater to the needs of low-achieving students in their classrooms. Cross-sectional assessments have consistently shown that most children cannot perform basic tasks in math and reading by the end of primary school (*1, 2*), revealing that they do not learn much while at school (*3*). Longitudinal studies have found that students' learning levels remain flat during primary and secondary school (*4-6*), suggesting that gaps in students' knowledge are rarely remedied during school. Classroom observations have documented that teachers often use the same materials for all students and spend most of their lesson time using whole-classroom instructional approaches, indicating that teachers rarely engage in differentiated or one-on-one instruction (*7, 8*).

The prevailing explanations for why teachers do not address the needs of their lowest-performing students focus on the incentives that they face. National curricula in developing countries are overly ambitious, and they encourage teachers to race to cover it at the expense of ensuring all students understand the material (*9*). Parents from low-income families often pull their children out of school if they do not show promise towards graduation, leading teachers to conclude that there is little payoff from investing in these children (*10, 11*). Teachers are informally appraised based on their students' performance on high-stakes exams, so they focus on the students with higher chances of taking and passing these exams (*12*).

In this paper, we argue that teachers pay insufficient attention to low-performing students partly because they underestimate how many of them there are in their classrooms and how much they are struggling. We leverage data on student achievement in math and reading (i.e., their actual scores on standardized tests) and teachers' estimates of student achievement (i.e., the scores that their own teachers expected them to obtain) from primary- and middle-school grades in India and Bangladesh and show that most teachers: (a) underestimate the *share* of low-scoring students in their own classrooms; (b) overestimate the *scores* of these students by a large margin (in absolute terms and in relative terms vis-à-vis the within-classroom achievement distribution); and (c) vastly underestimate within-classroom variation in achievement (i.e., test-score variability is an order of magnitude higher than teachers' estimations would imply).

We document similar trends across two contexts (India and Bangladesh), subjects (math and language), and types of assessment (one developed by an independent research team and one constructed by a government agency as part of a regular large-scale assessment program). We also rule out several potential confounders, including teachers not knowing who their students are (we ask them to recognize both real and fake students, and they are more likely to recognize the former) or not knowing how to translate their estimations into scores (we ask them to estimate student performance on specific topics and exercises and they are no more accurate at these levels; consistent with our results, they overestimate performance on easier exercises).

We show that none of the traditional teacher-quality metrics—having a master's degree, completing pre-service training, or having more experience—explain why some teachers are better at estimating student achievement than others, consistent with evidence from developing countries on the weak association between these metrics and student achievement gains (*13-16*). We also demonstrate that teachers' misestimations are not due to them making biased judgments against students who are female, from low-income families, or from scheduled castes and tribes. Instead, teachers appear to be using students' intelligence as a proxy for achievement.

Our study contributes to multiple ongoing debates in education in the developing world. First, it complements prior work emphasizing the importance of incentives for the mismatch between students' skills and teachers' instructional levels by highlighting the role of teachers' misestimations of their own students' skills (*17, 18*). Second, it prompts us to reexamine the promise of interventions designed to address this mismatch (*19-21*). These interventions are believed to benefit students mainly by changing how teachers deliver instruction, but they may play an equally important role in helping teachers update their beliefs about students' skills.

Our work also provides a useful contrast to similar studies in developed countries. The take-away from that body of research is that teachers' judgments are fairly accurate (*22-24*). We are among the first to show teachers' beliefs are far less accurate in developing settings, suggesting that some of the long-standing dysfunctions of developing-country school systems (e.g., high teacher absence rates, low subject-matter expertise among teachers; see *25, 26-28*), combined with some of the disruptions caused by recent expansions in access to schooling (e.g., increased heterogeneity in student preparation, larger class sizes; see *5, 12, 29*) may adversely impact instruction in these settings in more ways than previously acknowledged.

Results: ***Most teachers misestimate the share of low- and high-achieving students in their classroom.*** Teachers believe that there are fewer low-scoring students and more high-scoring students in their class than there actually are. In Bangladesh, we can compare the proportion of students in each classroom who scored in the bottom and top terciles in the 2019 national assessment to the proportion of students in that tercile estimated by teachers. The average teacher underestimates the share of bottom-tercile students in math by about 4 percentage points (pp.) and he/she overestimates the share of top-tercile students by almost 3 pp. (Figure 1 and Table S1 in the supplementary materials). We observe a similar pattern in language (Figure S1).

Teachers misestimate the performance of students across the achievement distribution. Teachers may not need to know their students' standing in the national achievement distribution, but they should be able to estimate their students' test performance on a subject that they teach. To examine whether they do, we calculate the difference between each student's score on each test (from 0 to 100) and his/her teacher's estimation of that score (from 0 to 100).

Teachers tend to overestimate the achievement of students in their classrooms in math. When we plot the differences between students' test-performance and teachers' estimations, we find that 84% of the differences in India and 59% of those in Bangladesh are positive, indicating that most teachers overestimate students' achievement (Figure 2). In India, the average estimation was 24 pp. higher than the average score; in Bangladesh, it was 8.5 pp. higher (Table S2). In Bangladesh, where we also collected data on students' performance and teachers' estimations in language (Bangla), most differences between performance and estimations are negative, suggesting that the direction of teachers' errors may differ across subjects (Figure S2).

Teachers did not make errors because they misunderstood what was covered in each test. In India, we used a test designed by an independent research team, but we showed it to teachers right before we elicited their estimations; in Bangladesh, we used a test administered as part of the national assessment, with which teachers were already familiar. Also, if teachers did not understand what was covered in each assessment, there is no reason why they would consistently *over*-estimate their students' performance.

It is also impossible for teachers' misestimations to be due to "mean reversion" (i.e., students who obtained low scores due to negative measurement error and students who obtained high scores due to positive measurement error both converging towards the mean). In India and Bangladesh, teachers were asked to estimate their students' performance on a single test, not their change in performance between two tests, so teachers' misestimations cannot be attributed to initially lower- or higher-than-expected scores that reverted to the mean. In India, students' performance and teachers' estimations were measured within days of each other, making it extremely implausible that teachers' errors reflect their anticipation of mean reversion.

Teachers' misestimations cannot be explained by teachers having information about their students that is not reflected in the tests. We did not ask teachers to estimate their students' knowledge or skills; we specifically asked them to estimate their students' percentage-correct scores on a test that they were either shown (in India) or that they knew (in Bangladesh). This is also why their misestimations cannot be explained by teachers overweighting the skills that they deem important when predicting students' scores, since the allocation of items to content and cognitive skills was revealed when the test was shown to teachers or understood from before.

The magnitude of teachers' misestimations of the students' test performance is large. One way to understand the magnitude of teachers' errors in their estimations of students' test performance is to compare it to the variability in test performance in their own classrooms. If two teachers misestimate the performance of a student by the same percentage points, but one teacher has students that vary more in their performance, his/her errors are more consequential because he/she will be less able to distinguish between any two students in his/her class. We find that teachers' estimations are incorrect by a large margin. Teachers' estimations diverge from students' scores by 126% of the within-classroom standard deviation (SD) of test performance in India and by 202% of the within-class SD in Bangladesh.

Another way to make sense of the magnitude of teachers' misestimations is to calculate the share of variation in students' test scores explained by teachers' estimations. Teachers' estimations predict a very small share of variability in students' actual test scores. If we regress students' test scores on teachers' estimated test scores and compute the corresponding coefficient of determination, teachers' estimations explain only about 13% of students' performance in performance in India and less than 1% in Bangladesh in math (Figure 3). We observe a similar figure for language (Figure S5 and Table S3).

The typical teacher overestimates the test scores of low-achieving students and underestimates those of high achievers. Given that teachers in developing countries devote little attention to students who struggle with the material, and that we suspect that this pattern may be partly due to teachers being unaware of the prevalence and degree of underperformance in their classrooms, it seems worth asking whether teachers are more likely to underestimate the test performance of a low-scoring student than that of a high-scoring classmate. If so, our results would be consistent with our hypothesized consequences of teachers' misestimations (although neither of our studies is set up to estimate how misestimations influence instruction).

Consistent with our first set of results, in Bangladesh, when we calculate the difference between students' actual test scores and teachers' estimations of those test scores separately for each tercile of the national achievement distribution (Figure 1 and Table S5), and for each tercile of the within-classroom distribution (Figure S4), we find that teachers consistently overestimate the test scores of students in the lowest tercile and underestimate those of students in the highest tercile. We observe a similar pattern in language (Figure S3 and Table S5).

Teachers do not even know the relative standing of students in their own classroom. Teachers may not need to accurately estimate the score of each of their students to understand that some students are struggling more with the material than others and require further support. If a teacher understands the relative standing (i.e., ranking) of students in his/her classroom, he/she should still be able to provide additional scaffolding to the students who need it most. Yet, teachers do not seem to know which students fare better than others. If we regress students' actual within-class rank (using test scores) on his/her estimated within-classroom rank (using teachers' estimations), implied rankings predict only 25% of variation in actual rankings in India and 2% in Bangladesh in math (Figures 4 and S6 and Table S3).

Teachers vastly underestimate the degree of variability in achievement in their classes. Teachers may not need to know the test performance of each individual student to understand that some of their students could benefit from remedial and/or differentiated instruction. Imagine, for example, that a student scores a 0 on a test, and another student scores a 100 on that test, and their teacher predicts the first student to score a 100 and the second student to score a 0. That teacher would be completely incorrect in his/her estimations of each student's test score, but he/she would hold an accurate estimation of the variability in achievement in his/her class. This awareness could lead the teacher to devote time to reinforce basic concepts or procedures.

Consistent with our results showing that teachers underestimate the test performance of low achievers and overestimate that of high achievers, which suggests that teachers believe that these two groups perform more similarly than they actually do, we find that teachers vastly underestimate variability in student achievement in their own classroom. If we compare the actual within-class SD (using students' test scores) to the implied within-class SD (using teachers' estimations), we find that 54% of teachers in India and 72% of those in Bangladesh underestimate the heterogeneity in test performance of their own students. If we regress the actual within-class SD on the implied within-class SD, we find that implied SDs explain only about 1% of variability in actual SDs in India and less than 1% in Bangladesh (Figure 4). We observe a similar pattern for language in Bangladesh (Figure S7).

Teachers' estimations are no more accurate when they focus on specific topics (e.g., geometry), items (e.g., finding the missing angle on an isosceles triangle), and/or students.

One possible explanation for the pattern documented above is that teachers know their students' academic skills but misunderstand how those skills translate into percentage-correct test scores.

Teachers' estimations, however, are no more accurate when they focus on specific topics. In Bangladesh, we asked teachers to estimate the share of students in their classrooms who are "proficient" (which we defined as "able to answer all test items on that topic") on five topics in the math test: operations, measurement, data, algebra, and geometry. Teachers overestimate the share of proficient students in *all* topics, by between 23 pp. (measurement and data) and 44 pp. (geometry; Figure 5 and Table S6). We also did the same for the three topics of the language test: reading, vocabulary, and grammar. In this case, teachers overestimate proficiency in reading but underestimate it for vocabulary and grammar (Figure S8 and Table S6).

Teachers' estimations are only slightly more accurate when they focus on specific items. In Bangladesh, we showed teachers specific questions in the test and asked them what percentage of their own students would be able to answer those questions correctly. The gaps between the estimated and actual shares were small in some cases (e.g., for the algebra question, the average difference was only 4.6 pp.), but it was still large in others (e.g., for the operations question, the average difference was more than 53 pp.; Figures 6 and S9 and Table S7).

The variability in teachers' predictions seems to be partly explained by item difficulty. In Bangladesh, we leverage the fact that we observe item-level performance for the items for which we asked teachers to estimate students' performance not only for their own students, but for the entire nationally representative sample of the country's sixth graders, and fit a two-parameter logistic Item Response Theory (IRT) model to estimate each item's difficulty. We find that teachers are particularly prone to over-estimate their students' performance on the most difficult items on the test in *both* math and language (see *b*-parameters in the x-axis of Figures 6 and S9). This explains why teachers overestimated students' performance on the reading item (which was the most difficult) but underestimated it for the other two items (which were easier).

Teachers' estimations are no more accurate when they focus on specific items *and* students. In Bangladesh, we asked teachers whether students in their own classrooms would be able to answer certain questions in the math test. We selected students from the bottom, medium, and top terciles of the national achievement and the within-classroom distributions. Consistent with our earlier results, teachers overestimate the performance of low achievers and they underestimate the performance of high achievers (Figures S10 and S11 and Table S8).

Teachers' misestimations are not due to them not knowing who their students are. In Bangladesh, we presented each teacher with the names of 12 students and asked them to verify whether they taught them. Of those 12 names, two were fake to check whether teachers claimed to recognize students who were not in their classroom. Teachers were far more likely to recognize the 10 real names (the average teacher recognized nearly 75% of these names) than the two fake names (the average teacher claimed to recognize only 30% of these names, and just 14% of the teachers indicated knowing both fake names). Further, teachers were just as likely to recognize students in the bottom and top terciles of the achievement distribution (Table S9).

Teachers who are more educated, trained, or experienced are no more likely to correctly estimate their students' test performance than their less well-prepared counterparts. None of the three measures commonly used to determine teachers' salaries—holding a master's degree, completing pre-service training, or having above-median experience (overall, at their school, or on the subject that they teach)—predicts which teachers estimate student achievement correctly. First, we regressed students' scores on teachers' estimations for teachers who have a given characteristic (e.g., master's) and those who do not, to explore whether the coefficient on the estimated scores differed in sign and magnitude across these two groups. Then, we regressed students' scores on teachers' estimation, an indicator variable for the characteristic, and the interaction of the last two variables, to examine whether differences were statistically significant. Yet, teachers' characteristics failed to predict the accuracy of estimations (Tables S10 to S14).

Teachers' misestimations are partly explained by their overreliance on heuristics. Teachers, like all individuals in general, seem to rely on heuristics when making estimations. Many teachers provide estimates around 60%, a widely used passing rate: in math, 11% of the estimates in India and 21% of those in Bangladesh were exactly at 60%, and 17% of those in India and 29% of those in Bangladesh were within 5 pp. of 60%. We observe a similar pattern among language teachers in Bangladesh. The use of these heuristics, however, does not explain teachers' misestimations. We obtain similar results if we run the regressions described above omitting teachers who offered estimates at or near 60% (Table S4).

Teachers' misestimations are more often explained by their overreliance on students' intelligence as a proxy for their test performance. The misestimations presented above are not explained by teachers consistently underestimating the scores of students with certain observable characteristics such as being female, from a low-income family, or from a scheduled caste. In India, where we have data on the characteristics of test-takers, we find that teachers overestimate the performance of all three of these groups, and teachers' estimates are no more accurate for students with these characteristics than for those without (Tables S15 to S17).

Instead, teachers seem to be over-weighting their subjective appraisals of each student's intelligence when estimating his/her achievement. If we classify each student based on whether he/she performed below or above his/her within-class average on a test of "fluid intelligence" (i.e., non-verbal abstract reasoning), we find that the mean score of the first group in the math test is 32% and that of the second group is 43%. Teachers' estimates of the mean scores of these two groups, however, are 56% and 70%, respectively. Thus, while the test scores of these groups differ by 11 pp., teachers' estimates differ by 14 pp., suggesting that teachers seem to be overweighting the importance of fluid intelligence (Table S18).

Discussion: Our study is one of the first to systematically document teachers' misestimations of student achievement in developing countries and to highlight its potential consequences for instruction. A recent review indicates that studies in high-income countries have found strong and positive correlations between teacher estimations of student achievement and students' actual achievement—the mean correlation cited in this review was $r = 0.63$ (*30*). We find the correlation between these two measures to be about *ten times weaker* in Bangladesh ($r = 0.07$) and *two times weaker* in India ($r = 0.36$).

Our results have three main implications for research. First, they draw attention to an arguably underappreciated reason why interventions that encourage teachers to assess the skills of their students, divide students into groups based on their performance, and assign each group to different activities (known as "differentiated instruction" or "teaching at the right level"), have been successful in South Asia, and more recently, in Sub-Saharan Africa (*19-21, 31, 32*). Given that teachers in these settings do not actually know the skills of their own students, a non-trivial part of the reason why these interventions have been effective is that they help teachers update their incorrect priors on the level and spread of achievement in their classrooms.

Second, our results suggest that providing teachers with the results of formative assessments or training them on how to administer such assessments may confer some of the benefits of differentiated instruction if teachers know how to adjust their instructional practices. If a meaningful share of the benefits of such programs stem from aligning teachers' estimations with students' performance (rather than from the ready-made activities for students at different levels that are often provided in differentiated-instruction interventions), correcting teachers' incorrect estimations may be a cost-effective approach to boost students' learning. Previous experiments with these strategies in lower-middle income countries have been disappointing (*33*), but they have been successful in upper-middle income economies (*34, 35*), suggesting that there is relatively little margin to raise teacher effort in contexts of high teacher absence (*25, 26, 36*). This approach may be worth trying in sub-national systems with higher attendance.

Third, our results raise important questions about the consequences of the documented prevalence of biases against disadvantaged students (e.g., girls, students from low-income families, or those from scheduled castes and tribes). It has already been shown that parents, teachers, and students often discriminate against these students (*37, 38*). What has not yet been examined is why and how these biased beliefs matter for students' daily lives at school. Future work ought to thus explore how teachers' misperceptions relate to their instructional practices; particularly, how teachers interact with students whom they under-estimate, and how these dyadic relationships affect their apparent reluctance to cater to the needs of low performers.

References

1. ASER. (ASER Centre, New Delhi, India, 2018).
2. Uwezo. (Uwezo East Africa Regional Office, Nairobi, Kenya, 2016).
3. World Bank, *World development report 2018: Learning to realize education's promise*. (The World Bank, Washington, DC, 2018).
4. T. Andrabi, J. Das, A. I. Khwaja, T. Vishwanath, T. Zajonc, *Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate*. (World Bank, Washington, DC, 2007).
5. K. Muralidharan, A. Singh, A. J. Ganimian, Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review* **109**, 1426-1460 (2019).
6. A. Singh, Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, 1-44 (2019).
7. S. Bhattacharjea, W. Wadhwa, R. Banerji. (ASER, Delhi, India, 2011).
8. D. Sankar, T. Linden, in *(South Asia Human Development Sector Report No. 67)*. (The World Bank, Washington, DC, 2014).
9. L. Pritchett, A. Beatty, Slow down, you're going too fast: Matching curricula to student skill levels. *International Journal of Educational Development* **40**, 276-288 (2015).
10. A. V. Banerjee, E. Duflo, in *Poor economics: A radical rethinking of the way to fight global poverty*. (Public Affairs, Philadelphia, PA, 2011).
11. S. Sabarwal, M. Abu-Jawdeh, R. Kapoor, Teacher beliefs: Why they matter and what they are. *The World Bank Research Observer* **37**, 73-106 (2022).
12. E. Duflo, P. Dupas, M. Kremer, Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review* **101**, 1739-1774 (2011).
13. M. Azam, G. G. Kingdon, Assessing teacher quality in India. *Journal of Development Economics* **117**, 74-83 (2015).
14. N. Bau, J. Das, in *Unpublished manuscript*. (The World Bank, Washington, DC, 2017).
15. J. Buhl-Wiggers, J. T. Kerwin, J. A. Smith, R. Thornton, in *Unpublished manuscript*. (Department of Economics, Copenhagen Business School, Copenhagen, Denmark, 2019).
16. M. C. Araujo, P. Carneiro, Y. Cruz-Aguayo, N. Schady, Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics* **131**, 1415-1453 (2016).

17. A. V. Banerjee, P. Glewwe, S. Powers, M. Wasserman, in *Unpublished manuscript*. (Abdul Latif Jameel Poverty Action Lab (J-PAL), Cambridge, MA, 2013).
18. P. Glewwe, K. Muralidharan, in *Handbook of economics of education*. (2016).
19. A. V. Banerjee *et al.*, From proof to concept to scalable policies: Challenges and solutions, with an application *Journal of Economic Perspectives* **31**, 73-102 (2017).
20. A. V. Banerjee, S. Cole, E. Duflo, L. Linden, Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics* **122**, 1235-1264 (2007).
21. A. V. Banerjee, R. Banerji, E. Duflo, R. Glennerster, S. Khemani, Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy* **2**, 1-30 (2010).
22. A. B. Feinberg, E. S. Shapiro, Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *Journal of Educational Research* **102**, 453–462 (2009).
23. J. C. Begeny, H. E. Krouse, K. G. Brown, C. M. Mann, Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review* **40**, 23-38 (2011).
24. M. K. Demaray, S. N. Elliot, Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly* **13**, 8 (1998).
25. N. Chaudhury, J. Hammer, M. Kremer, K. Muralidharan, F. H. Rogers, Missing in action: Teacher and health worker absence in developing countries. *The Journal of Economic Perspectives* **20**, 91-116 (2006).
26. K. Muralidharan, J. Das, A. Holla, A. Mohpal, The fiscal cost of weak governance: Evidence from teacher absence in India. *Journal of Public Economics* **145**, 116-135 (2017).
27. J. Metzler, L. Woessmann, The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics* **99**, 486-496 (2012).
28. T. Bold *et al.*, Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa. *Journal of Economic Perspectives* **31**, 185-204 (2017).
29. A. M. Lucas, I. M. Mbiti, Access, sorting, and achievement: The short-run effects of free primary education in Kenya. *American Economic Journal: Applied Economics* **4**, 226-253 (2012).
30. D. Urhahne, L. Wijnia, A review on the accuracy of teacher judgments. *Educational Research Review*, 100374 (2020).
31. A. V. Banerjee, R. Banerji, E. Duflo, M. Walton, in *Unpublished manuscript*. (2011).
32. S. A. Beg, A. R. Fitzpatrick, A. M. Lucas, in *Unpublished manuscript*. (University of Delaware, Newark, DE, 2021).
33. K. Muralidharan, V. Sundararaman, The impact of diagnostic feedback to teachers on student learning: Experimental evidence from India. *The Economic Journal* **120**, F187-F203 (2010).
34. R. de Hoyos, V. A. García-Moreno, H. A. Patrinos, The impact of an accountability intervention with diagnostic feedback: Evidence from Mexico. *Economics of Education Review* **58**, 123-140 (2017).
35. R. de Hoyos, A. J. Ganimian, P. A. Holland, Teaching with the test: Experimental evidence on diagnostic feedback and capacity-building for schools in Argentina. *World Bank Economic Review* **35**, 499-520 (2021).
36. M. Kremer, N. Chaudhury, F. H. Rogers, K. Muralidharan, J. Hammer, Teacher absence in India: A snapshot. *Journal of the European Economic Association* **3**, 658-667 (2005).
37. M. G. Farfan Bertran, A. Holla, R. Vakis, in *(Policy Research Working Paper No. 9593)*. (The World Bank, Washington, DC, 2021).
38. R. N. Hanna, L. L. Linden, Discrimination in grading. *American Economic Journal: Economic Policy* **4**, 146-168 (2012).

Acknowledgments: We thank Noam Angrist and T. M. Asaduzzaman for inputs in the design and analysis of the Bangladesh results; Rezarta Bilali, Erin Godfrey, Diane Hughes, Andy de Barros, Hiro Yoshikawa, Larry Aber, Mauricio Romero, and NYU seminar participants for comments; and Rashmi Menon for excellent research assistance in the original India study.

Funding: The study in India was funded by the Abdul Latif Jameel Poverty Action Lab's Post-Primary Education fund. The study in Bangladesh was funded by the World Bank and the China-World Bank Group Partnership Facility.

Author contributions:

Conceptualization: AJG, SS

Data curation: SD, AJG

Formal analysis: SD, AJG

Funding acquisition: AJG, SS

Methodology: AJG, SS, SD

Visualization: SD, AJG

Writing – original draft: AJG, SD

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data used in the Bangladesh study are obtained under the terms of a research collaboration agreement between the World Bank and the Monitoring and Evaluation Wing (MEW), Ministry of Secondary and Higher Education, Bangladesh. All data from India are obtained under terms of a memorandum of understanding between the Abdul Latif Jameel Poverty Action Lab's office for South Asia and the Department of Education, Pune Municipal Corporation (PMC), India. All data from India and Bangladesh data are available upon the establishment of a materials transfers agreement with MEW in Bangladesh and PMC in India.

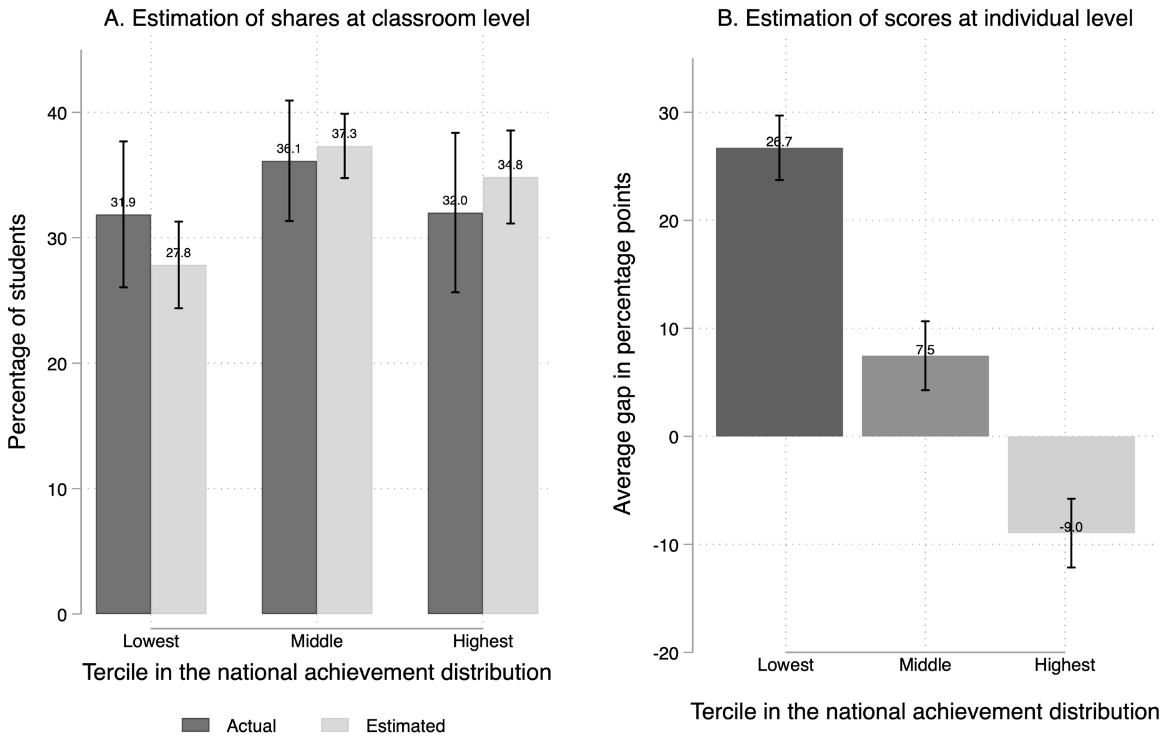

Fig. 1. Inaccuracies in teachers' estimations of students' academic skills at the classroom and individual level. In panel A, the figure shows the actual and estimated percentages of students at each tercile of the national achievement distribution on the 2019 National Assessment of Secondary Students (NASS) for math in Bangladesh, as indicated by students' scores (dark-gray bars) and teachers' estimations of those scores (light-gray bars). In panel B, the figure shows the differences between students' actual scores and teachers' estimations of those scores on the NASS for math in Bangladesh, by tercile of the national achievement distribution. In both panels, the error bars indicate 95% confidence intervals around the mean of each variable.

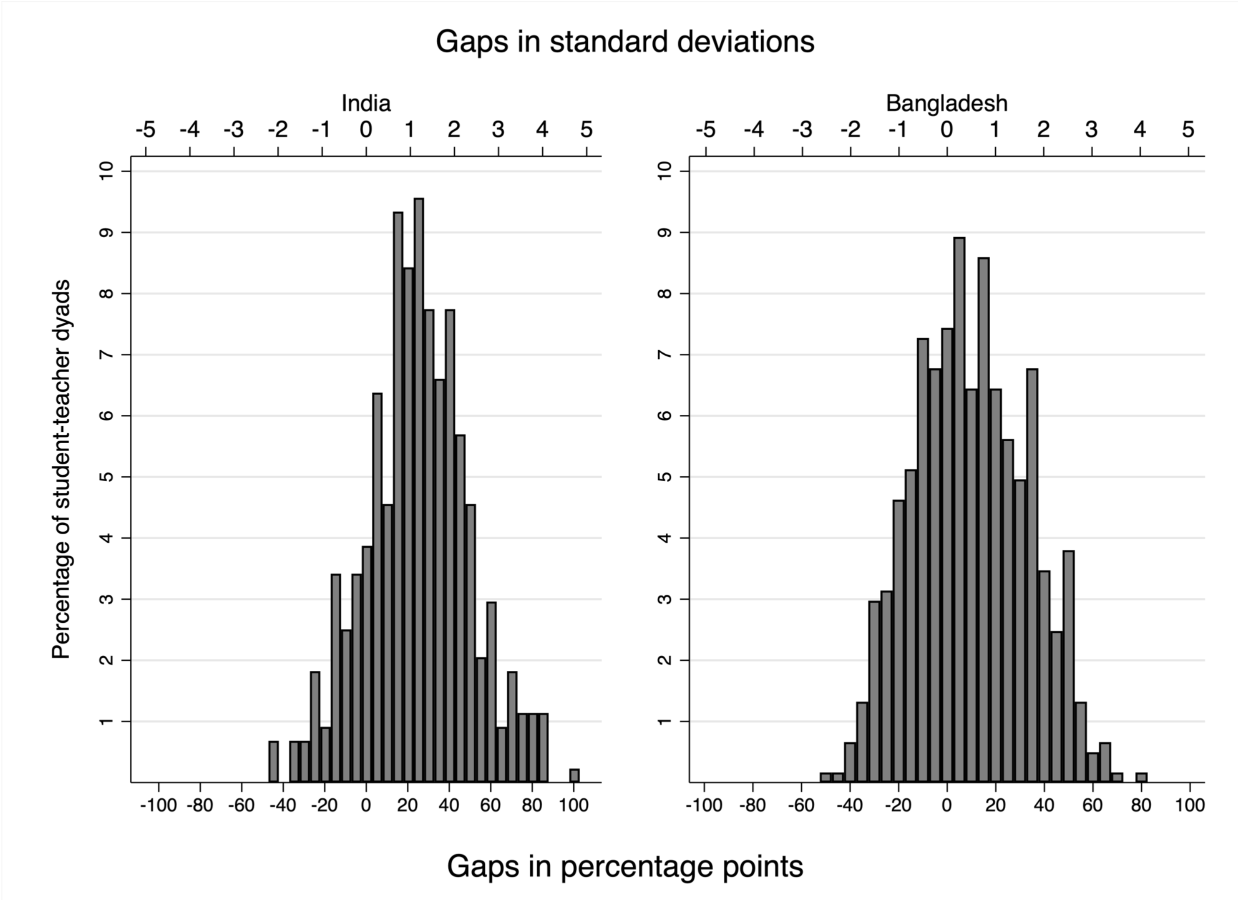

Fig. 2. Distribution of differences between actual and estimated percentage-correct scores. The figure shows distribution of differences between students' actual scores and teachers' estimations of those scores for math in India and Bangladesh, expressed in percentage points (bottom x-axis) and standard deviations of the within-class distribution (top x-axis). Observations are at the student-teacher dyad level.

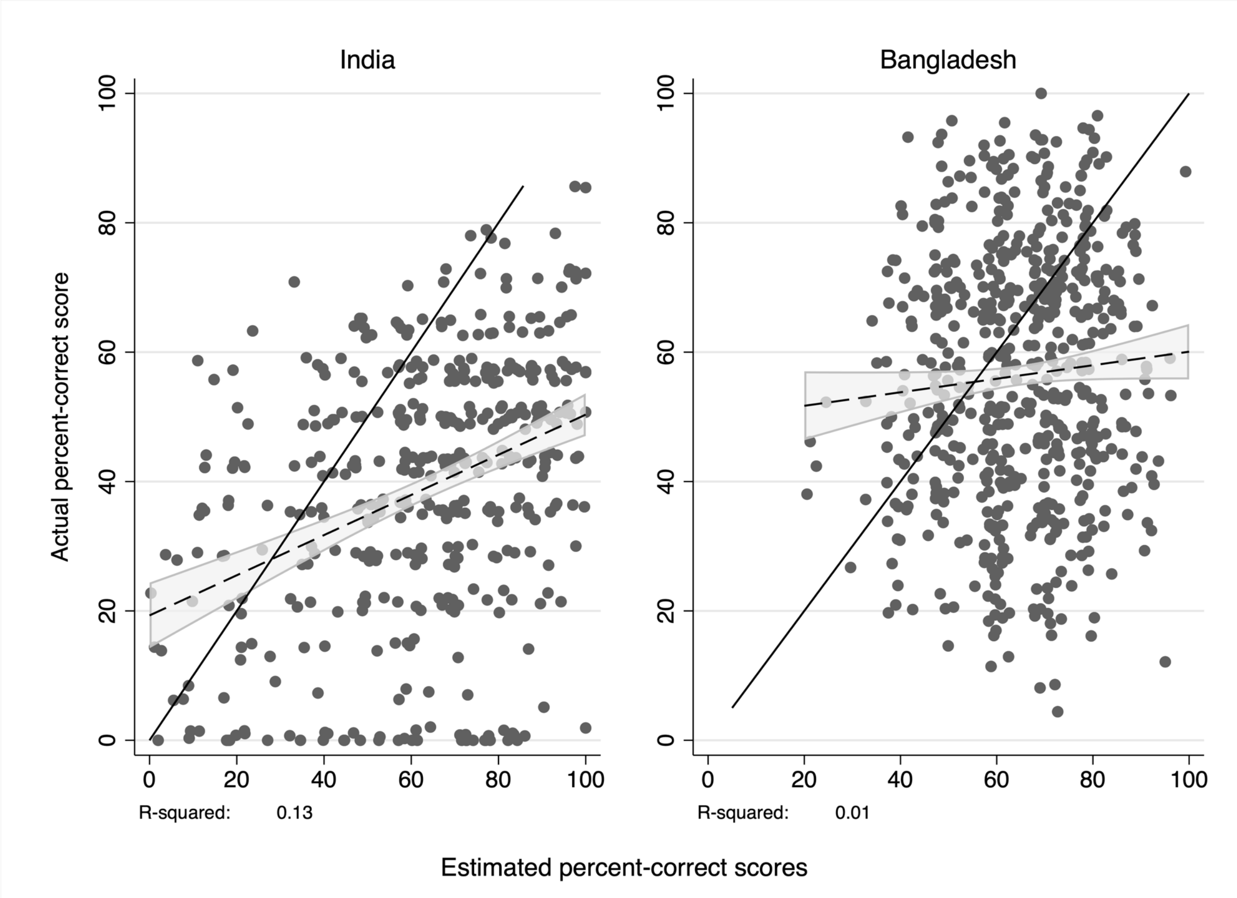

Fig. 3. Relationship between actual and estimated percentage-correct scores. The figure displays the relationship between students' actual scores and teachers' estimations of those scores for math in India and Bangladesh. Observations are at the student-teacher dyad level. The solid line indicates perfect prediction and the dotted line is the best-fit line with a 95% confidence interval (in gray).

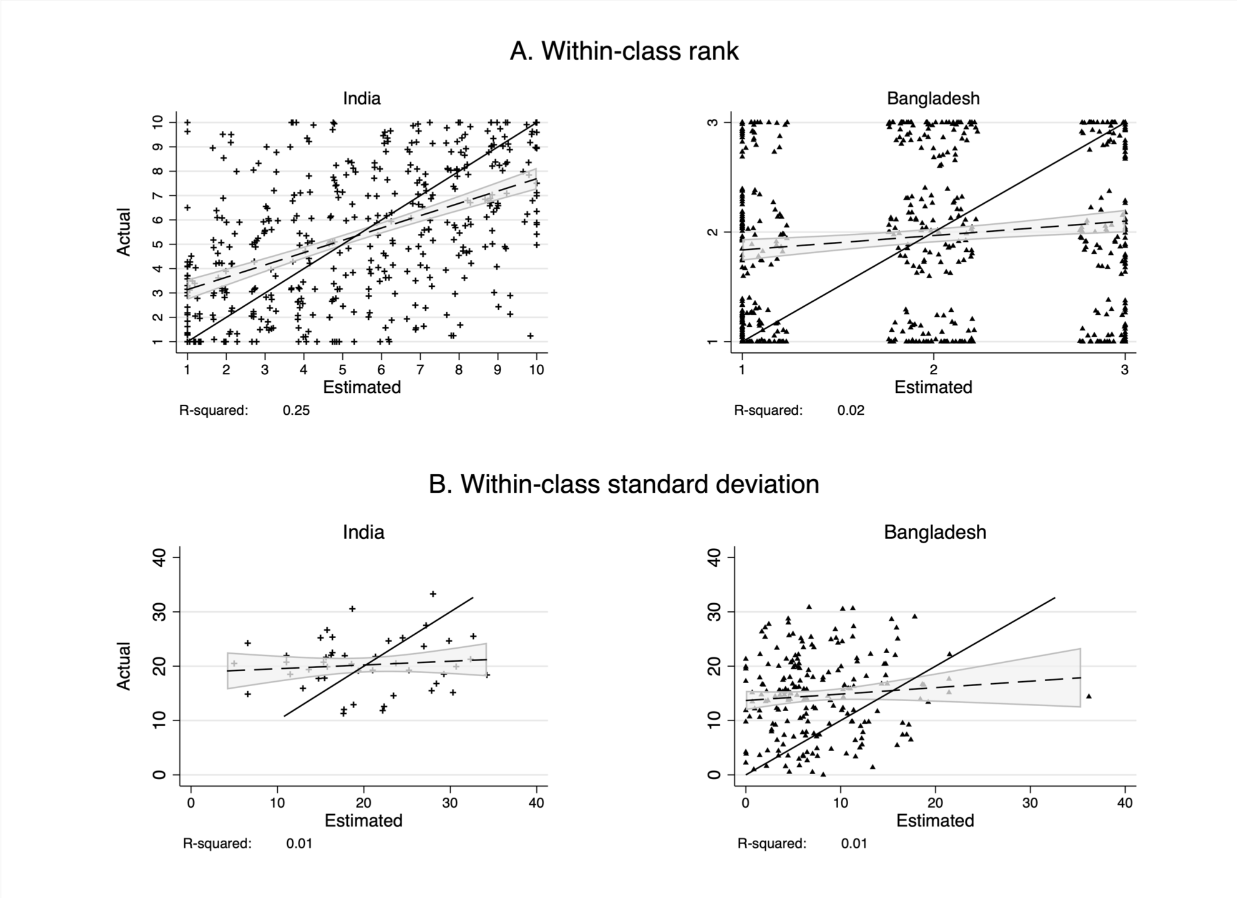

Fig. 4. Relationship between students' actual and estimated within-classroom rank and within-class standard deviation. In panel A, the figure displays the relationship between students' actual within-class ranks (based on their scores) and teachers' implied within-class ranks (based on their estimations of students' scores) for math in India and Bangladesh. The relationship is plotted for the three students matched with each teacher. A spherical random noise was added to the data before plotting to demonstrate the density of the discrete variable. In panel B, the figure displays the relationship between students' actual within-class standard deviations (based on their scores) and teachers' implied standard deviations (based on their estimations of students' scores) for math in Bangladesh. The solid line indicates perfect prediction and the dotted line is the best-fit line with a 95% confidence interval (in gray).

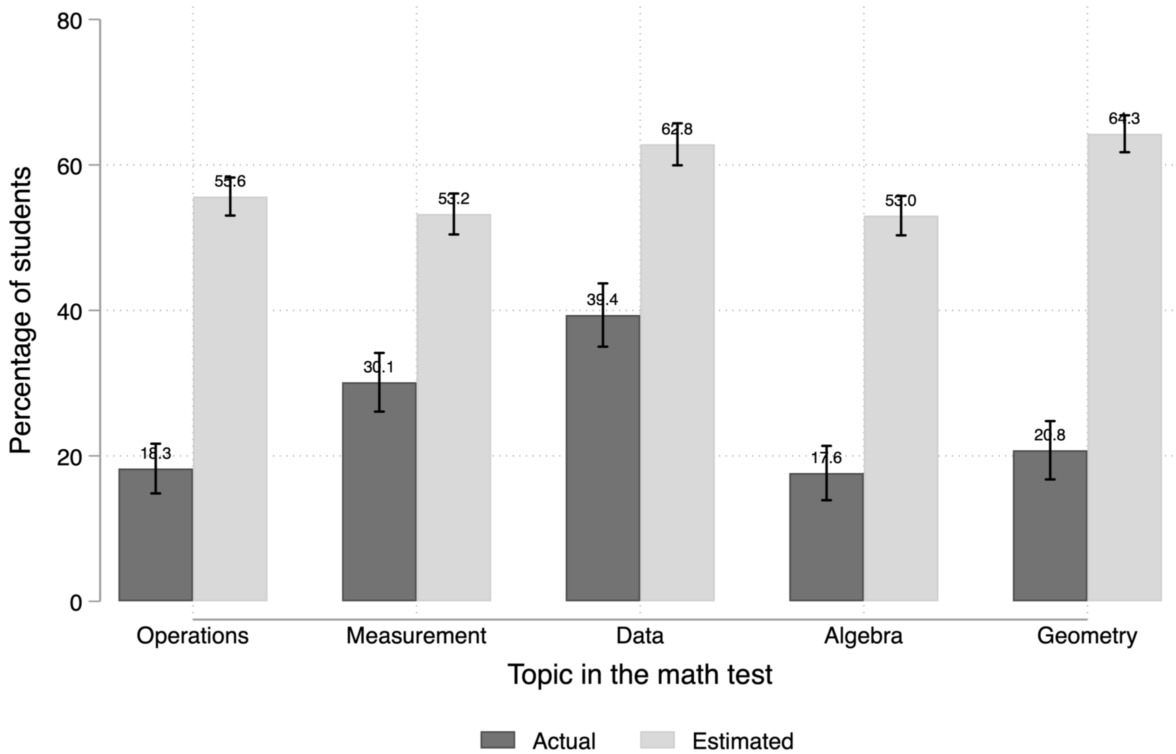

Fig. 5. Actual and estimated percentages of students who are proficient in each topic of the math test. The figure shows the actual and estimated percentages of students who are proficient (defined as being able to answer all the questions correctly) in each topic of the 2019 National Assessment of Secondary Students for math in Bangladesh, as indicated by students' scores (dark-gray bars) and teachers' estimations of those scores (light-gray bars). Error bars indicate 95% confidence intervals around the mean of each variable.

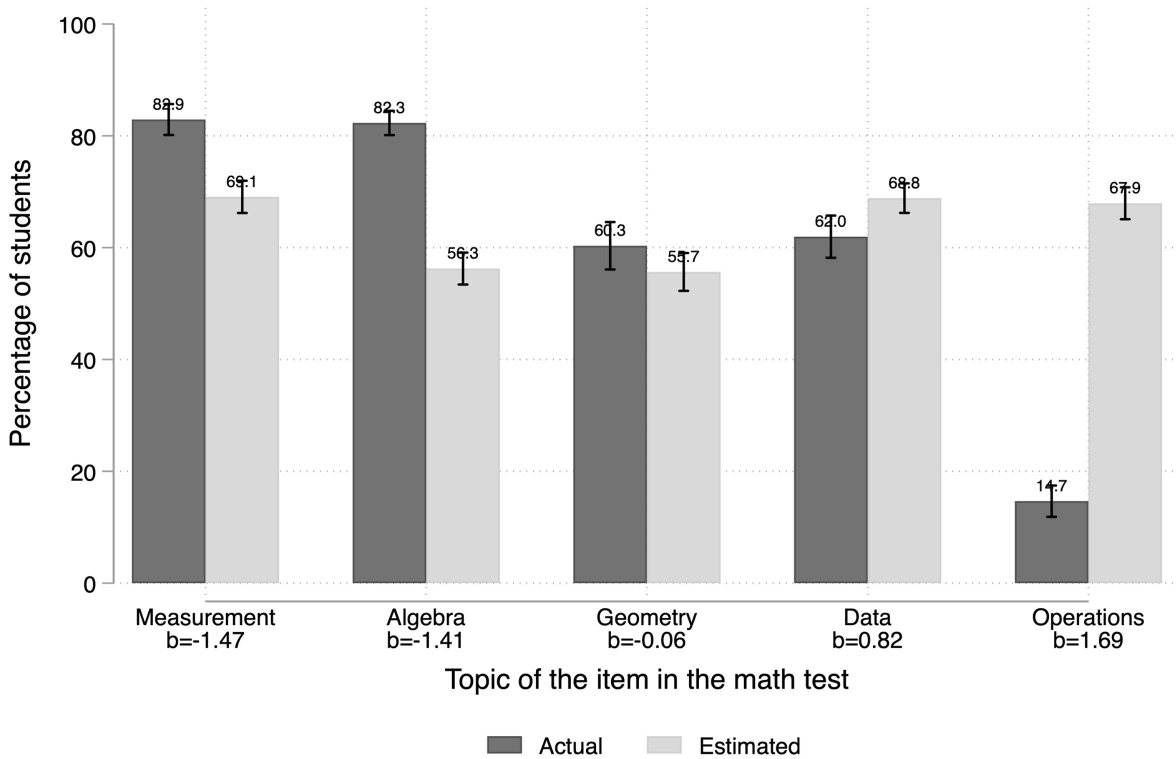

Fig. 6. Actual and estimated percentages of students who can answer a specific item in the math test in Bangladesh. The figure shows the actual and estimated percentages of students who can answer correctly a specific item on the 2019 National Assessment of Secondary Students for language in Bangladesh. Error bars indicate 95% confidence intervals around the mean of each variable. Teachers were asked to report their best estimation of the percentage of their students they think will answer the following questions correctly in each topic covered in the test: (a) the student was asked to arrange 4 numbers on a number line according to their value (e.g., 3, -7, 6, 0); (b) the student was given the length and the breadth of a rectangle and asked to calculate the area (e.g., 12 and 5 centimeters were given as length and the breadth); (c) the student was asked to find the mean and median of the five numbers (e.g., 9, 11, 6, 15, 11); (d) the student was asked how many bases there are in a math book; and (e) the student was asked to identify the x's coefficient in (-5x-5y). The difficulty (*b*) parameters are estimated using a two-parameter logistic item-response theory models with the full NASS sample, which is why they do not necessarily correspond to the percentage of students in our sample answering each item correctly.

Science
AAAS

Supplementary Materials for

Primary- and middle-school teachers in South Asia overestimate the performance of their students

Sharnic Djaker, Alejandro J. Ganimian, Shwetlena Sabarwal

Correspondence to: sharnic.djaker@nyu.edu

This PDF file includes:

Materials and Methods
Figs. S.1 to S.11
Tables S.1 to S.18

Materials and Methods

The materials and methods for both studies included in this paper were reviewed and approved by the Institutional Review Boards at the Institute for Financial Management and Research (for the study conducted in India) and the Institute of Health Economics at the University of Dhaka (for the one in Bangladesh).

In study 1, we use data from a randomized evaluation of a teacher-residency program in the state of Maharashtra, India. This evaluation was conducted in grades 5 and 6, in a convenience sample of 48 English-medium public primary schools in the city of Pune, in the Indian state of Maharashtra. The sample was selected as follows. The sampling frame included all 286 primary schools run by the Pune School Board (PSB). The authors of the evaluation then excluded 118 schools in remote rural areas (because it would have been challenging to monitor the intervention in those contexts), 30 Urdu-medium schools (because most of the teaching residents did not speak Urdu), 46 English-medium schools and 13 “model” schools that were implementing other interventions (to avoid confounding the effects of the teacher residency with these other programs), 20 schools with low enrollment (to minimize sampling error), and 9 schools that had already implemented the teaching residency. Shortly after the baseline, two schools had to be dropped due to a shortage of residents to staff them. We only use data from only the control group (to eliminate the possibility that the intervention may play a role in the patterns we observe) for regular teachers (i.e., not residents) for the endline round of data collection (when the researchers contemporaneously measured students’ achievement and teachers’ estimations). The subset of data that we use for this study includes 46 teachers and 457 students from grade 5 and 6.

In this study, we mainly use data from two instruments: a survey of teachers (https://bit.ly/3gCgMDe), which included questions eliciting their beliefs on the math skills of a random sample of 10 of their students; and student assessments (https://bit.ly/3xzNYRS), which evaluated the math skills of those students. Before making their estimations, teachers were shown 14 randomly chosen items from the assessments. The assessments included 35 multiple-choice items across three content domains (numbers, geometry and measures, and data display) and cognitive domains (knowing, applying, and reasoning). The distribution of items across content and cognitive domains was based on the assessment framework of the 2019 Trends in International Math and Science Study (TIMSS). The items covered a wide range of difficulty levels to minimize either zero or perfect scores. The survey of teachers and student assessments were administered within the same round of data collection.

In study 2, we use data from a study that we conducted in Bangladesh to understand the extent to which the patterns we observed in India were present in other school systems with similar institutional features. This study was conducted in grade 6, in a stratified random sample of 403 math and/or language (Bangla) teachers and their 1,306 students across 273 secondary public-private partnership (i.e., publicly funded, privately managed or PPP) schools. These schools account for more than 95% of enrollment in secondary education. We arrived at this sample as follows. First, we obtained access to the results of the 2019 National Assessment of Secondary Students (NASS)—a low-stakes large-scale assessment of a nationally representative sample of students in math, language, and English. This dataset included 28,238 students, whom we matched (using the exam cover sheets) to their 6,373 teachers across all three tested subjects. Then, we drew a simple random sample of three out of the eight divisions in the country (Chittagong, Dhaka, and Mymensingh) and we kept the 2,724 teachers in those divisions. Lastly, we excluded 1,034 teachers who did not teach either math or Bangla (the two target subjects in our study), 83 teachers for whom we did not have valid mobile numbers (which we needed to reach them), and 84 teachers for whom we did not have class-size information (which we needed for stratification). From the remaining 1,523 teachers, we randomly sampled 825 of them, stratifying our selection by a proxy for class size (whether schools were small, medium, or large, based on their grade 8 enrollment, since we could not obtain access to grade 6 enrollment figures). Of the sampled teachers, 573 were included in our target sample and 252 were allocated to a back-up roster. We called 726 teachers (573 teachers from the target sample and 153 from the back-up roster), we were able to connect with 607 of them, we obtained consent from 597 of them, we excluded 194 teachers (because they reported that they did not teach in the target schools, grades, or subjects), and we were ultimately left with 403 teachers. We offered each participating teacher a one-time cell phone credit of BDT 100 taka (~USD 1.17) to participate. We were able to link these teachers to 3,259 students, of whom teachers only recognized 2,445. We randomly selected 1,128 of these students to elicit teachers' estimations of their test performance.

In this study, we use data from two main sources: the 2019 NASS for grade 6, which allows us to observe students' skills in math and Bangla; and a phone-based survey of teachers that we developed (https://bit.ly/34phQYH) to elicit teachers' estimations of their students' test performance and collect additional information on potential moderators. The 2019 NASS included 40 items (of which 29 were multiple choice and 11 were open response) across five content domains in math (number and operations, measurement, data, geometry, and algebra) and 42 items (of which 36 were multiple choice and 6 open response) across three content domains in Bangla (reading, vocabulary, and grammar). The distribution of items across domains in both subjects was based on Bangladesh's national curriculum. It was administered in February and March of 2020 across the entire sample. In the survey of teachers, which we conducted in September of 2020, we first verified that respondents had taught math or Bangla to grade 6 students in the 2019 school year and asked them to identify a random subset of their students (for verification). Of the 403 teachers, 212 taught math, 181 taught Bangla, and 10 taught both. Teachers who taught both the subjects were randomly assigned into one of the two subjects for the survey and we asked teachers about their demographics, education, and experience. In the last part, we elicited teachers' estimations of their students' scores on the 2019 NASS.

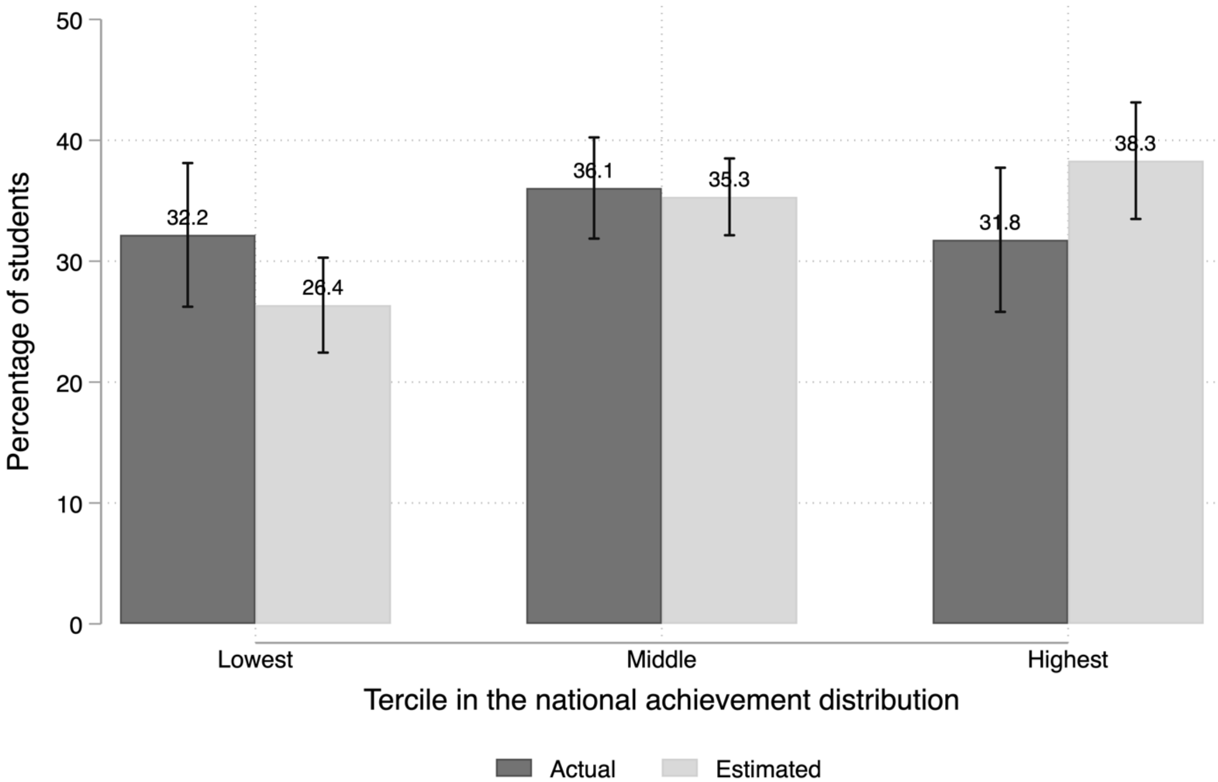

Figure S1. Actual and estimated percentages of students at each tercile of the national achievement distribution on the 2019 National Assessment of Secondary Students for language in Bangladesh, as indicated by students' scores (dark-gray bars) and teachers' estimations of those scores (light-gray bars). Error bars indicate 95% confidence intervals around the mean of each variable.

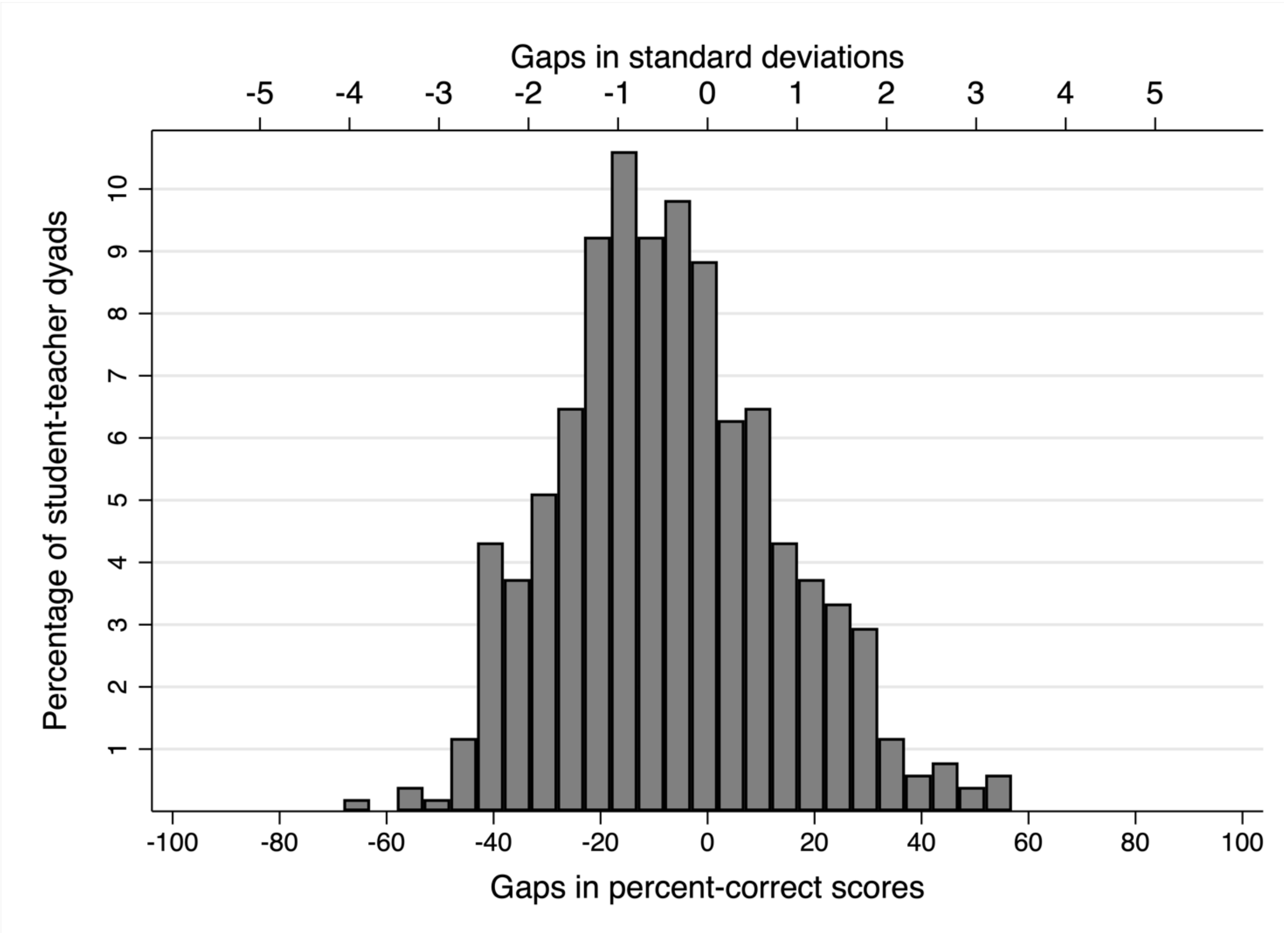

Figure S2. Distribution of differences between students' actual scores and teachers' estimations of those scores on the 2019 National Assessment of Secondary Students for language in Bangladesh, expressed in percentage points (bottom x-axis) and standard deviations of the within-class distribution (top x-axis). Observations are at the student-teacher dyad level.

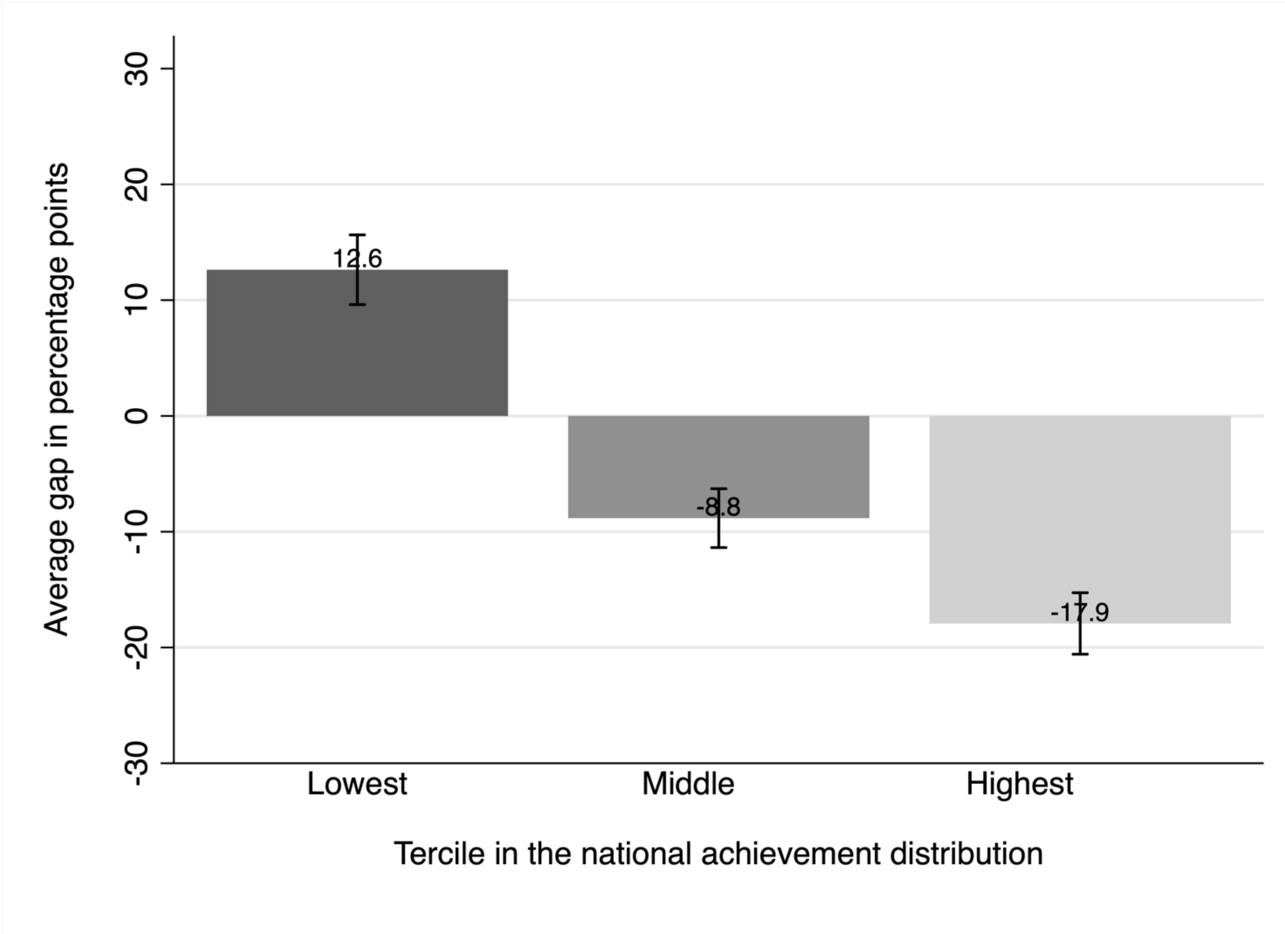

Figure S3. Differences between students' actual scores and teachers' estimations of those scores on the 2019 National Assessment of Secondary Students for language in Bangladesh, by tercile of the national achievement distribution. Error bars indicate 95% confidence intervals around the mean of each variable.

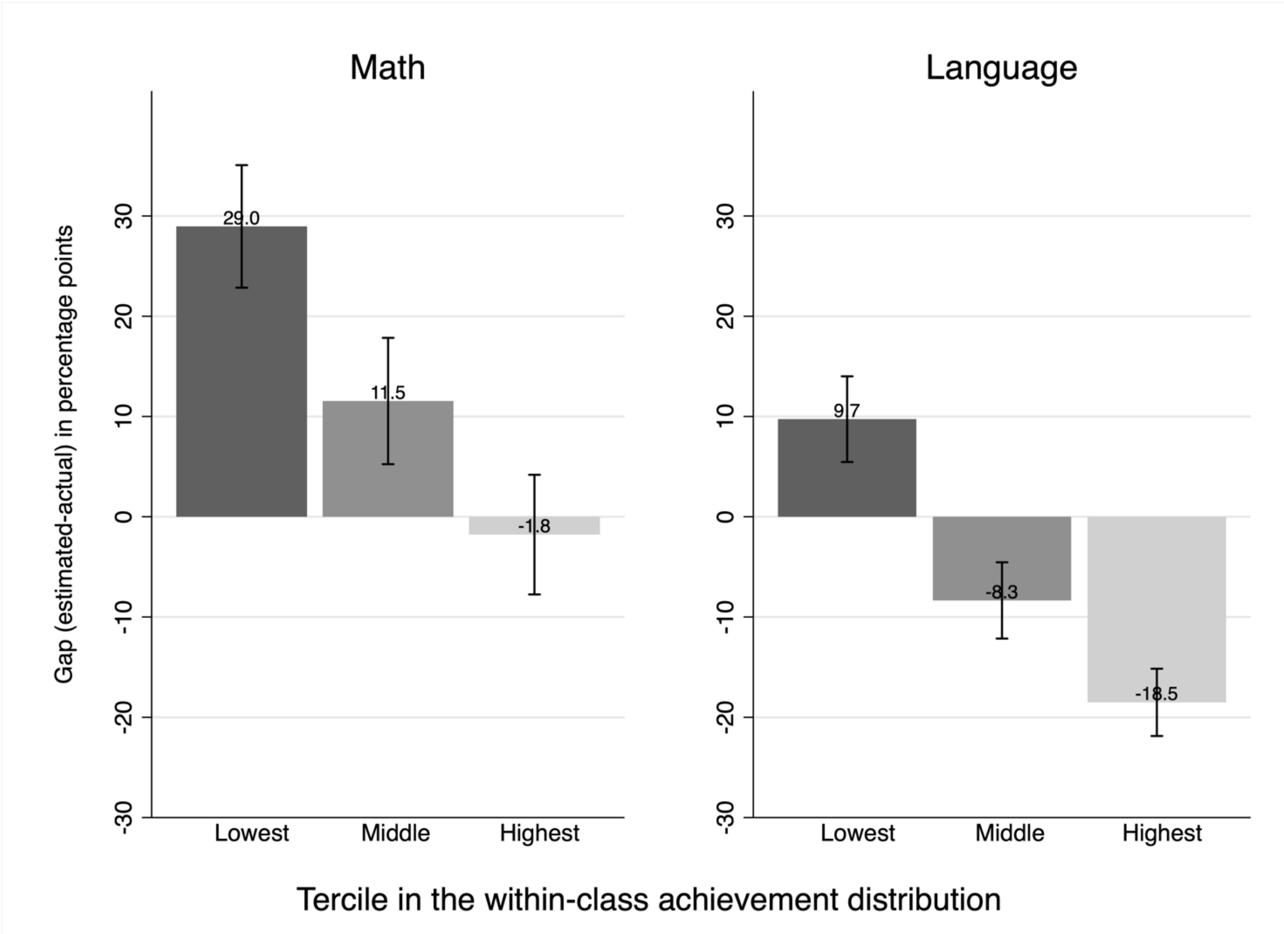

Figure S.4. Differences between students' actual scores and teachers' estimations of those scores on the 2019 National Assessment of Secondary Students for math (left panel) and language (right panel) in Bangladesh, by tercile of the within-class distribution. Error bars indicate 95% confidence intervals around the mean of each variable.

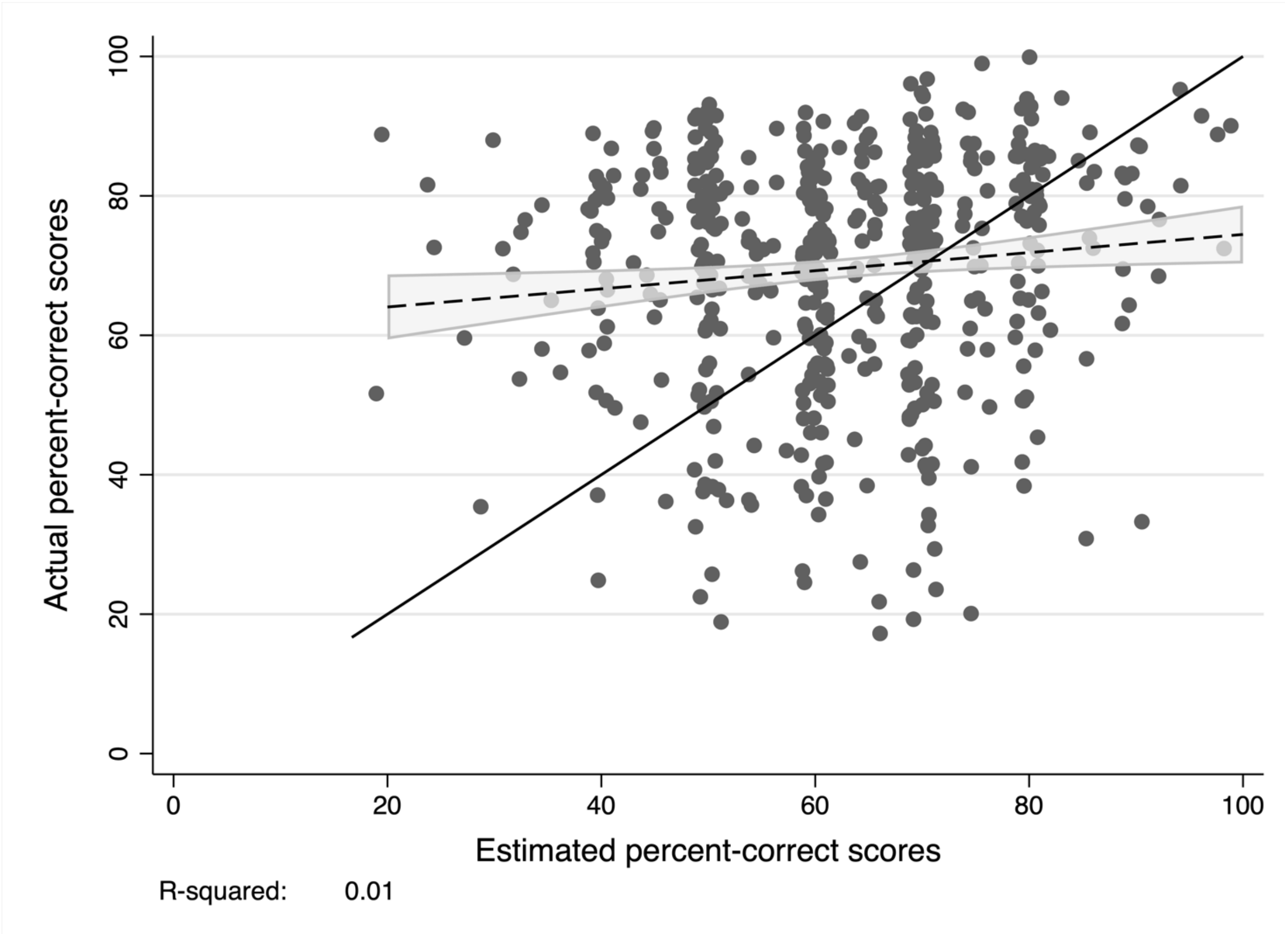

Figure S5. Relationship between students' actual scores and teachers' estimations of those scores on the 2019 National Assessment of Secondary Students for language in Bangladesh. Observations are at the student-teacher dyad level. The solid line indicates perfect prediction and the dotted line is the best-fit line with a 95% confidence interval (in gray).

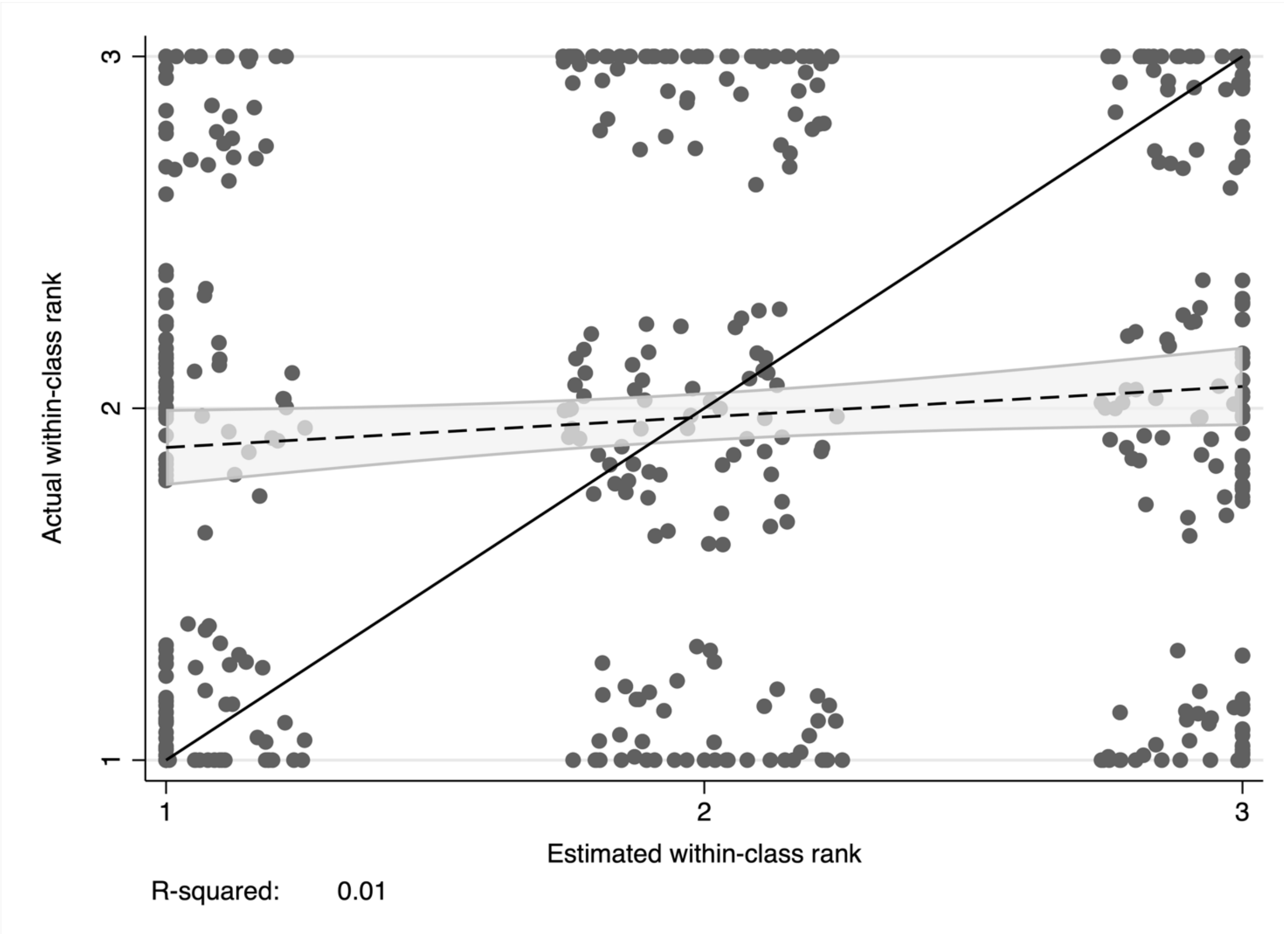

Figure S6. Relationship between students' actual within-class ranks (based on their scores) and teachers' implied within-class ranks (based on their estimations of students' scores) on the 2019 National Assessment of Secondary Students for language in Bangladesh. The relationship is plotted for the three students matched with each teacher. A spherical random noise was added to the data before plotting to demonstrate the density of the discrete variable. The solid line indicates perfect prediction and the dotted line is the best-fit line with a 95% confidence interval (in gray).

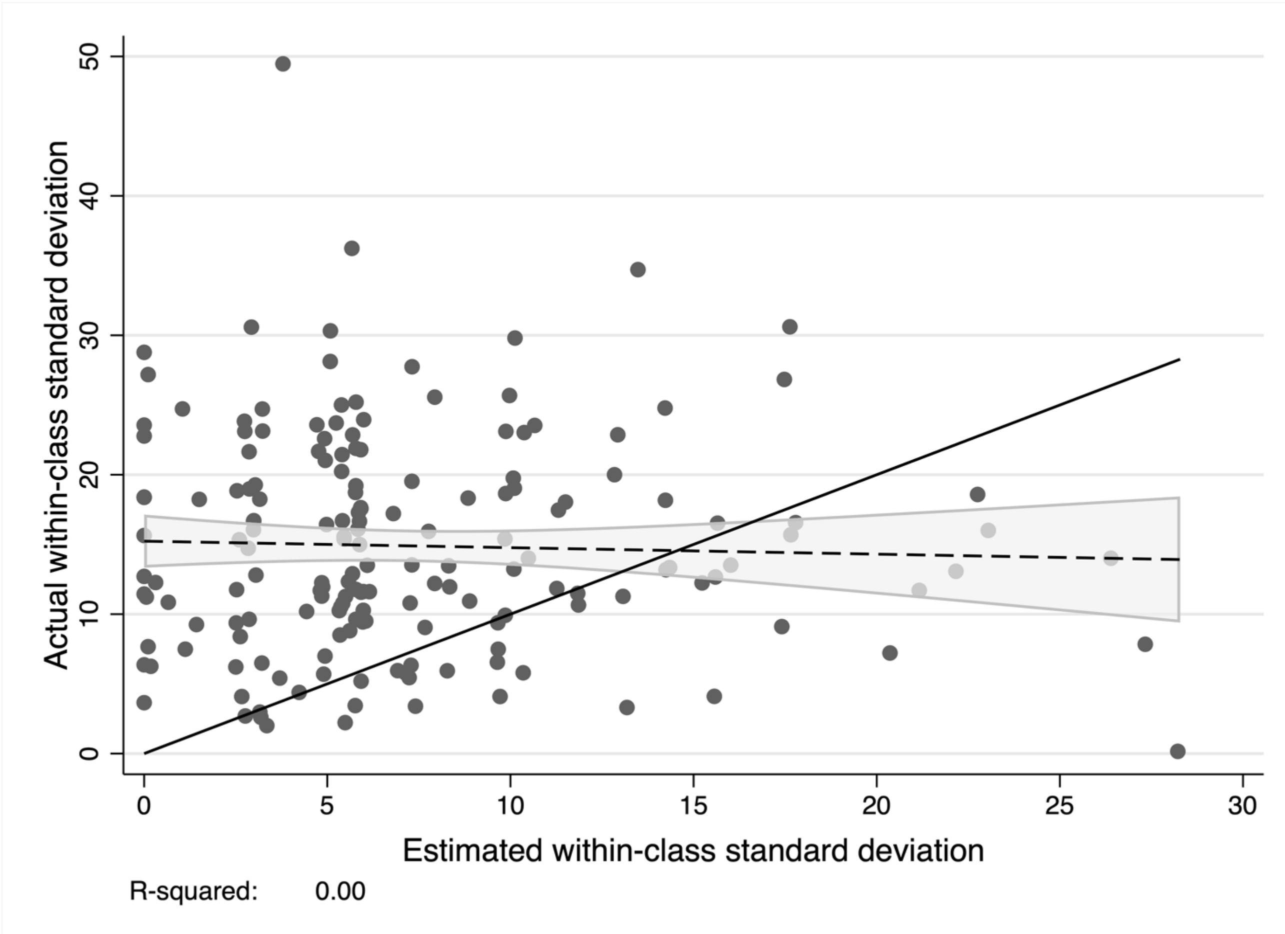

Figure S7. Relationship between students' actual within-class standard deviations (based on their scores) and teachers' implied standard deviations (based on their estimations of students' scores) on the 2019 National Assessment of Secondary Students for language in Bangladesh. The solid line indicates perfect prediction and the dotted line is the best-fit line with a 95% confidence interval (in gray).

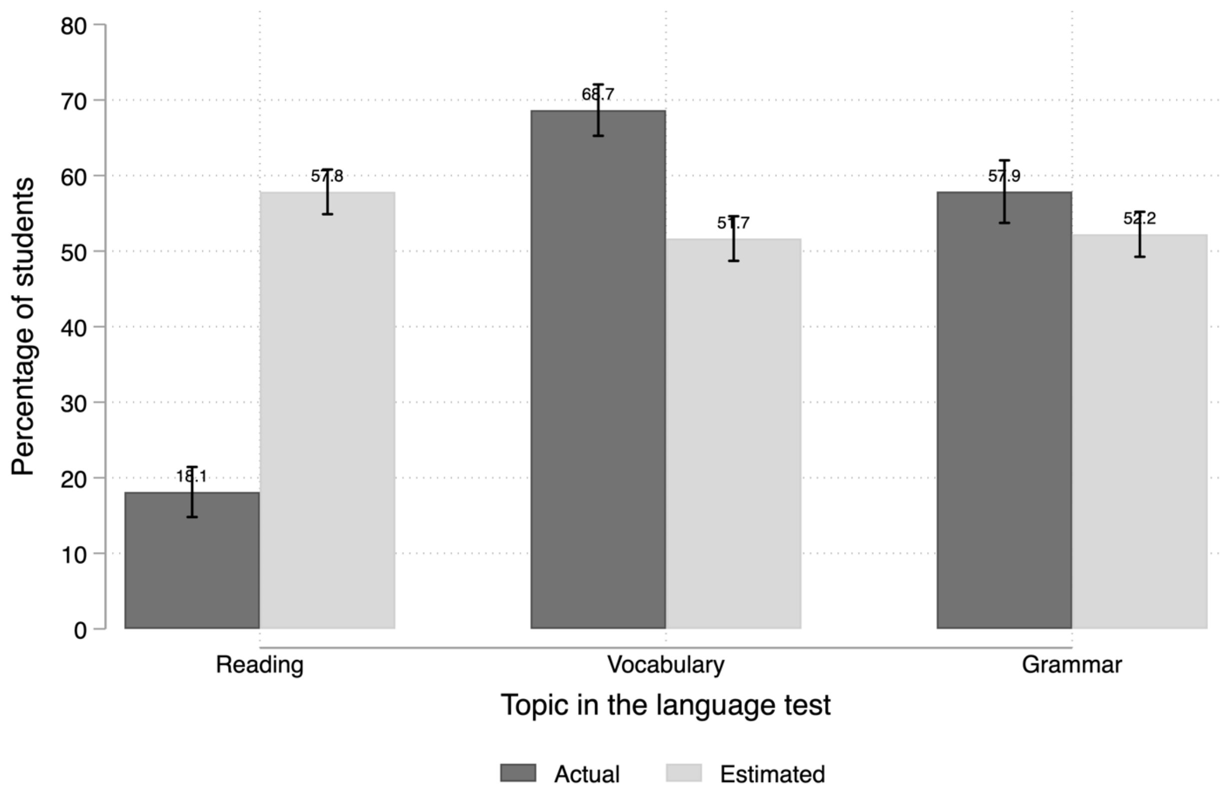

Figure S8. Actual and estimated percentages of students who are proficient (defined as being able to answer all the questions correctly) in each topic of the 2019 National Assessment of Secondary Students for language in Bangladesh, as indicated by students' scores (dark-gray bars) and teachers' estimations of those scores (light-gray bars). Error bars indicate 95% confidence intervals around the mean of each variable.

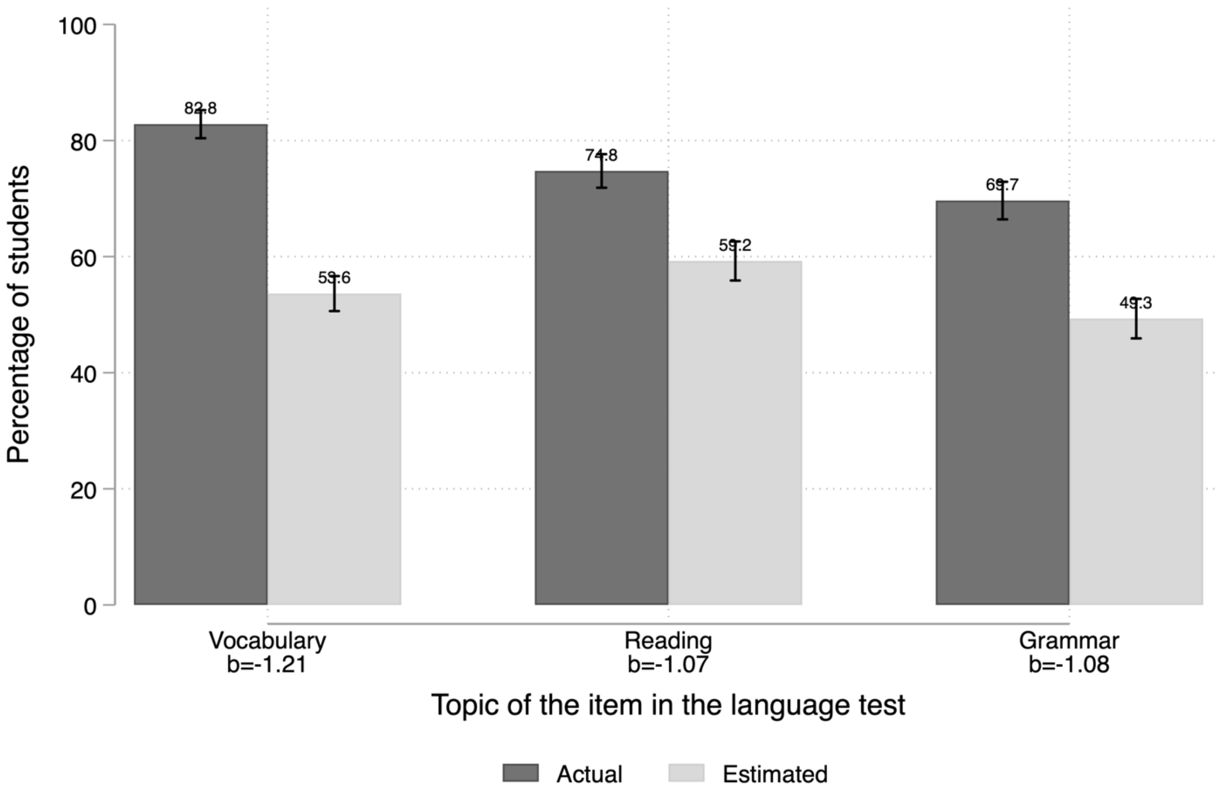

Figure S9. Actual and estimated percentages of students who can answer correctly a specific item on the 2019 National Assessment of Secondary Students for language in Bangladesh. Error bars indicate 95% confidence intervals around the mean of each variable. Teachers were asked to report their best estimation of the percentage of their students they think will answer the following questions correctly in each topic covered in the test: (a) What is the opposite of water? (b) What elements according to this paragraph are causing environmental pollution? (c) Which two letters combine to form the Bangla letter *khio*?

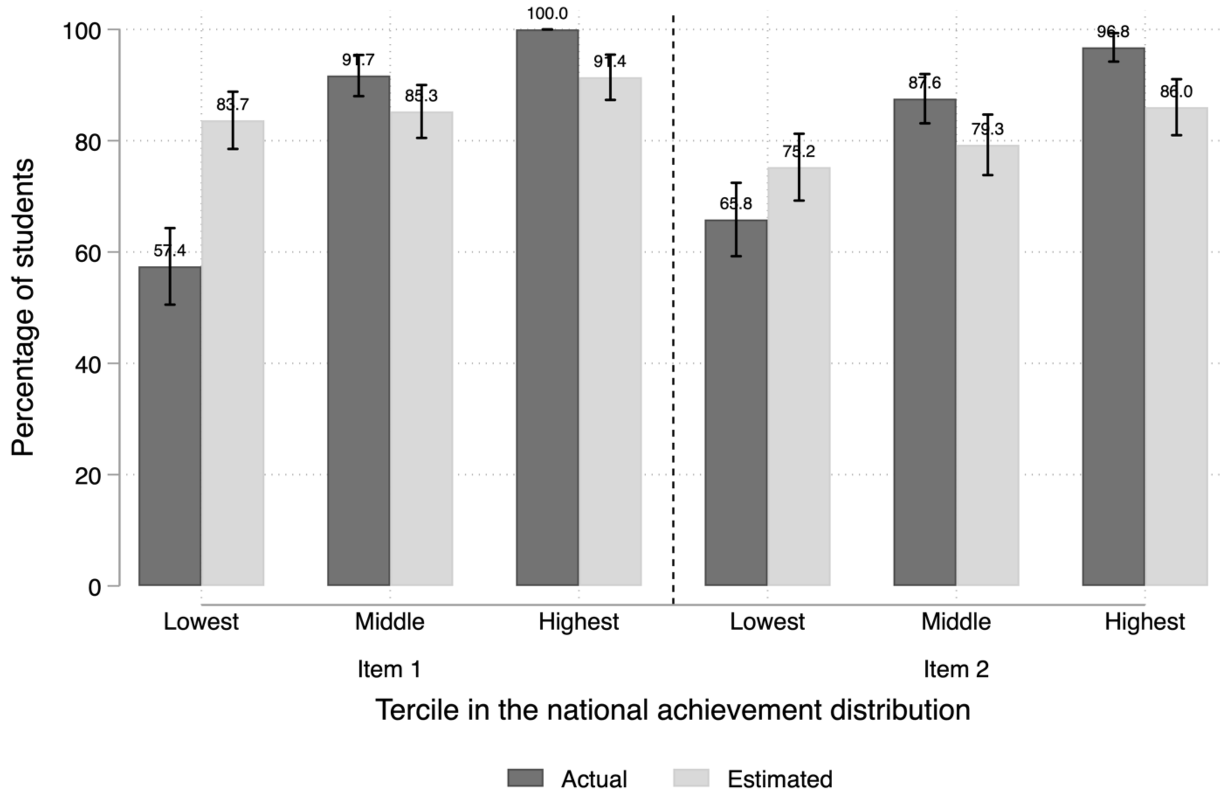

Figure S10. Actual and estimated percentages of students who can answer correctly two specific items on the 2019 National Assessment of Secondary Students for math in Bangladesh, by terciles of the national achievement distribution. Item 1 is "What is the area of the rectangle with length 12 and breadth 5 centimeters?" and item 2 is "What is the coefficient of x in −5x − 5y?" Error bars indicate 95% confidence intervals around the mean of each variable.

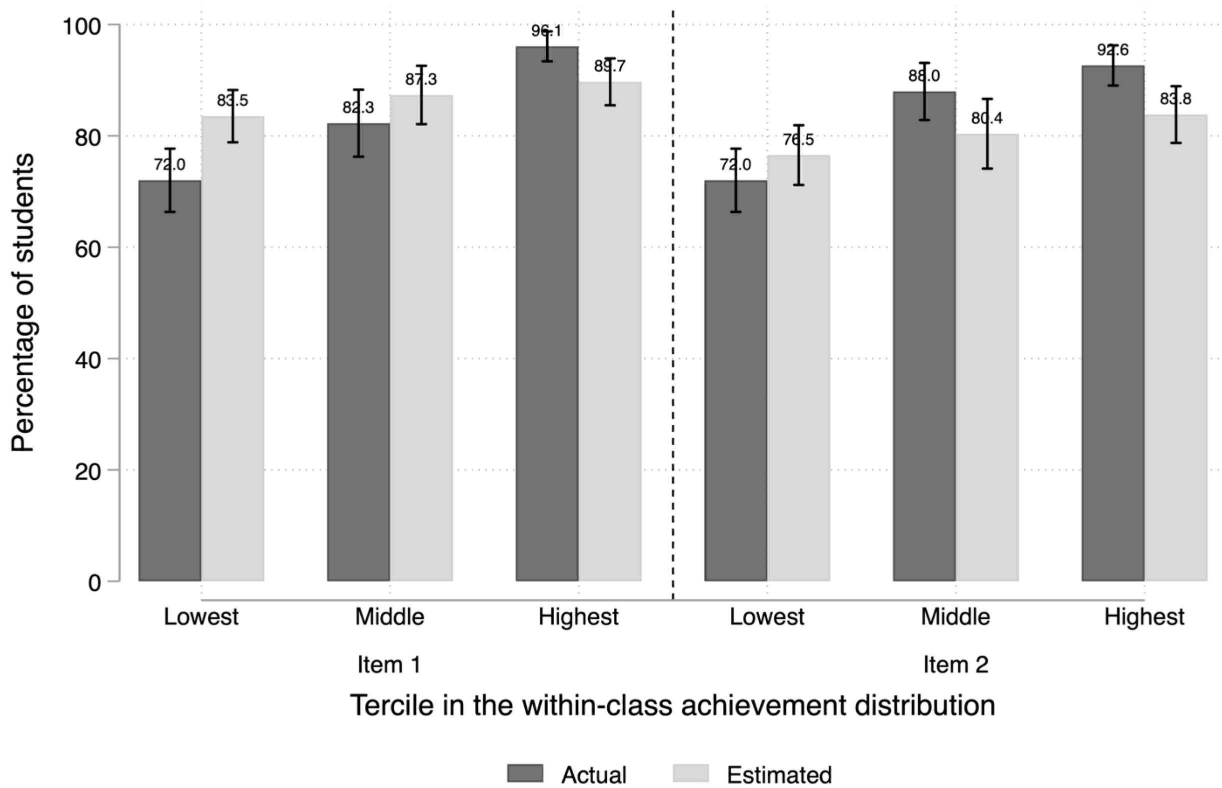

Figure S11. Actual and estimated percentages of students who can answer correctly two specific items correctly on the 2019 National Assessment of Secondary Students for math in Bangladesh, by within-class terciles of the achievement distribution. Item 1 is “What is the area of the rectangle with length 12 and breadth 5 centimeters?” and item 2 is “What is the coefficient of x in −5x − 5y?” Error bars indicate 95% confidence intervals around the mean of each variable.

Table S1. Actual and estimated percentages of students at each performance level of national achievement distribution in the math and language tests in Bangladesh

	(1)	(2)	(3)
A. Math	Estimated	Actual	Col. (1)-(2)
Percentage of students in the...			
...lowest tercile	27.83	31.86	-4.02
	[18.71]	[31.49]	(4.68)
...middle tercile	37.32	36.14	1.19
	[13.89]	[26.01]	(3.89)
...highest tercile	34.84	32.00	2.84
	[20.08]	[34.40]	(4.53)
N (teachers)	115	115	230
B. Language			
Percentage of students in the...			
...lowest tercile	26.37	32.18	-5.81
	[19.07]	[28.85]	(4.91)
...middle tercile	35.32	36.05	-0.73
	[15.43]	[20.32]	(4.10)
...highest tercile	38.31	31.77	6.54
	[23.40]	[28.95]	(5.21)
N (teachers)	93	93	186

Notes: The table shows the estimated and actual percentage of students in each tercile of the national achievement distribution. Column (1) shows teachers' estimations and column (2) shows actual percentages. Column (3) shows the difference between estimated and actual percentages (positive differences indicate over-estimations and negative differences indicate underestimations). Each observation corresponds to statistic for the classroom. The data is based on questions 26 and 36, for math and language respectively, in the questionnaire. Teachers were asked, "estimate what percent of students in your class are at the following performance levels based on their performance in NASS." This question was administered to 50% of the teachers in the survey selected randomly. Standard errors (clustered at the teacher level) are in parentheses and standard deviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S2. Distribution of the difference between actual and estimated percent-correct scores in India and Bangladesh

	(1) Mean	(2) Standard deviation	(3) Minimum	(4) Maximum	(5) 25th percentile	(6) 50th percentile	(7) 75th percentile	(8) N (student-teacher dyads)
A. India								
Math								
Difference between students' scores and teachers' estimations of those scores…								
… as percentage points	24.23	25.10	-47.14	100.00	8.57	24.29	40.00	439
…as share of the within-class standard deviation	-0.00	1.18	-3.36	3.57	-0.74	0.00	0.74	439
B. Bangladesh								
Math								
Difference between students' scores and teachers' estimations of those scores…								
… as percentage points	8.45	23.44	-52.50	82.50	-10.00	7.50	25.00	605
…as share of the within-class standard deviation	-0.00	1.18	-3.07	3.73	-0.93	-0.05	0.83	605
Language								
Difference between students' scores and teachers' estimations of those scores…								
… as percentage points	-7.02	20.58	-68.10	56.67	-20.95	-8.81	6.19	509
…as share of the within-class standard deviation	-0.00	1.23	-3.65	3.81	-0.83	-0.11	0.79	509

Notes: The table shows the summary statistics of the student-teacher level gaps between estimated and the actual percent-correct scores in India and Bangladesh. Differences are shown both in percentage points and as shares of the within-class standard deviation. Each observation corresponds to a teacher-student dyad. In both countries, teachers were asked to report their estimations of the percent-score that each student mentioned by name to him/her will score on a test (see Materials and Methods for details).

Table S3. Relationship between students' actual and estimated percent-correct scores in India and Bangladesh

	(1)	(2)
A. India	Actual score	Actual rank
Math		
Estimated score or rank	0.310***	0.506***
	(0.043)	(0.039)
N (student-teacher dyads)	439	439
R-squared	0.131	0.250
B. Bangladesh		
Math		
Estimated score or rank	0.105*	0.126**
	(0.061)	(0.051)
N (student-teacher dyads)	605	610
R-squared	0.005	0.016
Language		
Estimated score or rank	0.130**	0.166***
	(0.053)	(0.056)
N (student-teacher dyads)	509	518
R-squared	0.012	0.027

Notes: The table shows the results from the regression of actual percent-correct scores and within-class ranks on estimated percent-correct scores and ranks. Each cell is from a separate regression of actual measure on the estimated measured with and without teacher fixed effects. Each observation corresponds to a teacher-student dyad. Standard errors (clustered at the teacher level) are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S4. Relationship between students' actual and estimated percent-correct scores omitting teachers who offered anchored-estimates in India and Bangladesh

	(1)	(2)
A. India	Omitting teachers who estimated scores at 60%	Omitting teachers who estimated scores within 5 pp. of 60%
Math		
Estimated score	0.312***	0.312***
	(0.043)	(0.044)
N (student-teacher dyads)	389	363
R-squared	0.146	0.154
B. Bangladesh		
Math		
Estimated score	0.090	0.095
	(0.061)	(0.061)
N (student-teacher dyads)	477	429
R-squared	0.005	0.006
Language		
Estimated score	0.121**	0.114**
	(0.053)	(0.053)
N (student-teacher dyads)	412	348
R-squared	0.012	0.013

Notes: The table shows the results from the regression of actual and estimated percent-correct scores omitting teachers whose estimations were 60% or within 5 percentage points of 60%. Each cell is from a separate regression. Each observation corresponds to a teacher-student dyad. Standard errors (clustered at the teacher level) are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S5. Average actual and estimated percent-correct scores of students at each performance level of national achievement distribution in the math and language tests in Bangladesh

A. Math	(1) Estimated	(2) Actual	(3) Col. (1)-(2)
Average percent-correct score of students in the…			
…lowest tercile	63.19	36.47	26.72***
	[12.88]	[7.60]	(2.15)
…middle tercile	64.32	56.86	7.46***
	[14.23]	[6.93]	(2.30)
…highest tercile	67.27	76.22	-8.96***
	[13.33]	[6.50]	(2.29)
N (student-teacher dyads)	270	270	540
B. Language			
…lowest tercile	61.89	49.26	12.63***
	[12.07]	[10.75]	(2.05)
…middle tercile	63.03	71.86	-8.83***
	[12.30]	[4.85]	(1.73)
…highest tercile	66.55	84.48	-17.93***
	[13.16]	[3.68]	(1.81)
N (student-teacher dyads)	300	300	600

Notes: The table shows the average estimated and actual percent-correct scores for students in each tercile of the national achievement distribution. Column (1) shows teachers' estimations and column (2) shows actual percent-correct scores. Column (3) shows the difference between estimated and actual scores. Each observation corresponds to a teacher-student dyad. The table only includes those teachers who had reported an estimation for at least one student in each of the performance levels. Standard errors (clustered at the teacher level) are in parentheses and standard deviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S6. Actual and estimated percentages of students who are proficient in each content domain in math and language tests in Bangladesh.

	(1)	(2)	(3)
A. Math	Estimated	Actual	Col. (1)-(2)
Students can correctly answer all the questions in...			
Operations	55.65	18.26	37.39***
	[19.55]	[25.53]	(3.15)
Measurements	53.25	30.12	23.13***
	[21.01]	[30.14]	(3.60)
Data	62.83	39.36	23.47***
	[21.55]	[32.51]	(3.73)
Algebra	53.02	17.64	35.38***
	[20.21]	[27.93]	(3.29)
Geometry	64.28	20.77	43.51***
	[18.94]	[29.97]	(3.59)
N (teachers)	217	217	434
B. Language			
Reading	57.84	18.11	39.72***
	[20.43]	[22.91]	(3.05)
Vocabulary	51.66	68.65	-17.00***
	[20.40]	[23.48]	(3.21)
Grammar	52.22	57.87	-5.65
	[20.57]	[28.64]	(3.52)
N (teachers)	186	186	372

Notes: The table shows the estimated and actual percentage of students proficient in each content domain. A student is proficient in a domain if he/she gets all the questions from the domain correct. Column (1) shows teachers' estimations at the classroom level, and column (2) shows actual percentages. Column (3) shows the difference between estimated and actual percentages (positive differences indicate over-estimations and negative differences indicate underestimations). Each observation corresponds to a one classroom. The data is based on questions 29 and 39, for math and language respectively in the questionnaire. Teachers were asked to report the percent of students in their class who will answer all questions correctly in the given content area. Standard errors (clustered at the teacher level) are in parentheses and standarddeviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S7. Actual and estimated percentage of students who can answer a given question correctly in math and language tests in Bangladesh.

A. Math	(1) Estimated	(2) Actual	(3) Col. (1)-(2)
Students can correctly answer the shown question in...			
Operations	67.94	14.65	53.28***
	[21.32]	[20.95]	(2.87)
Measurements	69.07	82.91	-13.84***
	[21.56]	[20.67]	(2.86)
Data	68.85	61.96	6.89*
	[19.74]	[28.20]	(3.55)
Algebra	55.65	60.34	-4.69
	[25.34]	[31.67]	(4.05)
Geometry	56.25	82.30	-26.05***
	[21.32]	[16.28]	(2.63)
N (teachers)	217	217	434
B. Language			
Reading	59.25	74.75	-15.51***
	[23.23]	[19.95]	(3.07)
Vocabulary	53.63	82.82	-29.19***
	[20.84]	[16.74]	(2.75)
Grammar	49.32	69.65	-20.34***
	[23.53]	[22.37]	(3.14)
N (teachers)	186	186	372

Notes: The table shows the estimated and actual percentage of students who can answer a particular question from each of the content domains correctly. Column (1) shows teachers' estimations, and column (2) shows actual percentages. Column (2) shows the difference between estimated and actual percentages (positive differences indicate over-estimations and negative differences indicate underestimations). Each observation corresponds to a one classroom. The data is based on questions 31 and41, for math and language respectively in the questionnaire. Teachers were asked, "what percent of your students do you think can answer the following questions correctly?". Standard errors (clustered at the teacher level) are in parentheses and standard deviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S8. Actual and estimated percentage of students, grouped by tercile of the national achievement distribution, who can answer a given question correctly in Bangladesh.

	(1)	(2)	(3)	(4)	(5)	(6)
	Area of the rectangle with length 12 and breadth 5 centimeters			*Coefficient of x in −5x − 5y*		
A. Math	Estimated	Actual	Col. (1)-(2)	Estimated	Actual	Col. (4)-(5)
i. National distribution						
Low achievers	83.66	57.43	26.24***	75.25	65.84	9.41
	[37.06]	[49.57]	(5.96)	[43.26]	[47.54]	(6.02)
Medium achievers	85.25	91.71	-6.45	79.26	87.56	-8.29
	[35.54]	[27.64]	(3.96)	[40.64]	[33.08]	(5.11)
High achievers	91.40	100.00	-8.60***	86.02	96.77	-10.75***
	[28.12]	[0.00]	(2.58)	[34.77]	[17.72]	(3.74)
N (dyads)	605	605	1,210	605	605	1,210
ii. Within-class distribution						
Low achievers	83.54	72.02	11.52**	76.54	72.02	4.53
	[37.16]	[44.98]	(4.91)	[42.46]	[44.98]	(5.33)
Medium achievers	87.34	82.28	5.06	80.38	87.97	-7.59
	[33.36]	[38.31]	(5.26)	[39.84]	[32.63]	(5.19)
High achievers	89.71	96.08	-6.37*	83.82	92.65	-8.82**
	[30.46]	[19.46]	(3.25)	[36.91]	[26.16]	(4.37)
N (dyads)	605	605	1,210	605	605	1,210
B. Language	*Main theme of the reading comprehension*			*Animal that can live both in land and under water*		
i. National distribution						
Low achievers	84.56	45.64	38.93***	89.93	51.01	38.93***
	[36.25]	[49.98]	(6.85)	[30.19]	[50.16]	(7.04)
Medium achievers	88.89	82.01	6.88	89.95	82.01	7.94*
	[31.51]	[38.51]	(5.01)	[30.15]	[38.51]	(4.70)
High achievers	87.72	95.91	-8.19**	94.15	97.66	-3.51
	[32.92]	[19.87]	(3.93)	[23.53]	[15.16]	(2.90)
N (dyads)	509	509	1,018	509	509	1,018
ii. Within-class distribution						
Low achievers	86.29	58.88	27.41***	90.86	65.48	25.38***
	[34.48]	[49.33]	(5.84)	[28.89]	[47.66]	(5.45)
Medium achievers	86.62	77.46	9.15	88.73	78.87	9.86*
	[34.16]	[41.93]	(6.22)	[31.73]	[40.97]	(5.93)
High achievers	88.82	94.71	-5.88	94.12	92.35	1.76
	[31.60]	[22.46]	(3.97)	[23.60]	[26.65]	(3.43)
N (dyads)	509	509	1,018	509	509	1,018

Notes: The table shows the estimated and actual percentage of students grouped by terciles from the national achievement distribution who can answer the given question correctly. Low achievers scored in the bottom tercile (0-33) and high achievers scored in the top tercile (67-100) in the test. Columns (1) and (4) show the average of teachers' estimations at an individual student level; columns (2) and (5) show actual percentages of students who could answer. Columns (3) and (6) show the difference between estimated and actual percentages (positive differences indicate over-estimations and negative differences indicate underestimations). Teachers were shown two items for each student. Each observation corresponds to a student-teacher dyad. The data is based on questions 32 and 42 in the questionnaire. Teachers were asked, "can the students listed below answer the following questions correctly?". Standard errors (clustered at the teacher level) are in parentheses and standard deviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S9. Average percentage of student-names recognized by the teachers in each tercile of the national and within-class distribution in Bangladesh.

	(1)	(2)	(3)
	List of names restricted to		
A. National distribution	…1 per Level	…2 per level	No restrictions
Students grouped by...			
...all performance levels	75.72	76.56	74.50
	[23.33]	[22.50]	[23.97]
...low performance level	75.32	76.84	75.33
	[31.55]	[28.83]	[30.44]
...medium performance level	74.89	75.34	73.52
	[29.02]	[29.03]	[30.22]
...high performance level	78.22	77.64	76.20
	[32.42]	[29.56]	[31.52]
N (teachers)	233	145	403
B. Within-class distribution			
...all performance levels	74.15	73.34	74.50
	[23.17]	[22.44]	[23.97]
...low performance level	73.86	72.84	73.93
	[30.54]	[27.55]	[31.26]
...medium performance level	74.74	72.97	75.05
	[33.30]	[30.75]	[33.05]
...high performance level	75.20	74.31	75.18
	[32.37]	[30.98]	[32.50]
N (teachers)	336	195	403

Notes: The table shows the percentage of student-names recognized by teachers. The rows show estimates grouped by tercile of the national achievement distribution and the columns show the estimates grouped by the minimum number of names shown in each of the performance levels. Panel A shows students groups by performance in the nationals distribution while panel B is based on within-class distribution. Column (1) is restricted to teachers who had at least one student name shown from each performance level. Column (2) is restricted to teachers who had at least 2 student names shown from each level. Column (3) shows the average recognition when teachers are shown any number of children from each level. Each observation corresponds to a teacher. The data is based on question 15 in the questionnaire. Teachers were readout 10 names of students from their class asked to report whether they recognized this student.

Table S10. Predictive power of teacher's education level in India and Bangladesh.

	(1)	(2)	(3)	(4)	(5)	(6)
	With a master's degree or above		Without a master's degree		Test for interaction effects	
A. India						
Math						
Estimated	0.211***	0.375***	0.381***	0.496***	0.381***	0.496***
	(0.068)	(0.046)	(0.055)	(0.059)	(0.055)	(0.059)
Covariate					7.611	20.640***
					(6.538)	(4.883)
Interaction					-0.170*	-0.121
					(0.086)	(0.074)
Covariate + Interaction					7.441	20.519
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	166	166	273	273	439	439
R-squared	0.061	0.341	0.193	0.355	0.146	0.351
B. Bangladesh						
Math						
Estimated	0.104	0.277*	0.105	0.398**	0.105	0.398**
	(0.072)	(0.157)	(0.111)	(0.173)	(0.111)	(0.173)
Covariate					-0.094	-20.198
					(8.821)	(15.499)
Interaction					-0.001	-0.121
					(0.132)	(0.233)
Covariate + Interaction					-.094	-20.32
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	378	378	227	227	605	605
R-squared	0.005	0.572	0.005	0.564	0.005	0.569
Language						
Estimated	0.130*	0.489***	0.121	0.384***	0.121	0.384***
	(0.069)	(0.156)	(0.081)	(0.145)	(0.081)	(0.144)
Covariate					2.273	34.840***
					(6.791)	(11.347)
Interaction					0.009	0.106
					(0.106)	(0.212)
Covariate + Interaction					2.282	34.946
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	298	298	211	211	509	509
R-squared	0.013	0.386	0.009	0.361	0.019	0.380

Notes: The table shows degree of association between the teacher characteristic and their estimations. Columns (1)-(4) show results from the regression of actual percent-correct scores on the estimated separately for teachers with a postgraduate level education (master's degree or above), and for those with undergraduate level education or below. Columns (5) and (6) show the differences between both the groups. The data on teachers' education levels were collected from the survey of teachers in both India and Bangladesh. Each observation in the estimation is a teacher-student dyad. Standard errors (clustered at the teacher level) are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S11. Predictive power of teacher training status in India and Bangladesh.

	(1)	(2)	(3)	(4)	(5)	(6)
	With pre-service training		Without pre-service training		Test for interaction effects	
A. India						
Math						
Estimated	0.326***	0.457***	0.293***	0.437***	0.293***	0.437***
	(0.059)	(0.055)	(0.065)	(0.061)	(0.064)	(0.060)
Covariate					0.698	11.380**
					(6.255)	(4.628)
Interaction					0.033	0.020
					(0.087)	(0.081)
Covariate + Interaction					.731	11.4
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	266	266	173	173	439	439
R-squared	0.128	0.300	0.138	0.418	0.136	0.348
B. Bangladesh						
Math						
Estimated	0.114	0.335**	0.073	0.292	0.073	0.292
	(0.071)	(0.136)	(0.122)	(0.237)	(0.121)	(0.235)
Covariate					-1.739	25.687
					(9.681)	(16.921)
Interaction					0.042	0.043
					(0.140)	(0.272)
Covariate + Interaction					-1.697	25.73
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	461	461	144	144	605	605
R-squared	0.007	0.578	0.002	0.535	0.006	0.569
Language						
Estimated	0.142**	0.465***	0.096	0.393**	0.096	0.393***
	(0.067)	(0.143)	(0.082)	(0.151)	(0.081)	(0.150)
Covariate					-1.273	-44.159***
					(6.803)	(11.153)
Interaction					0.046	0.072
					(0.105)	(0.207)
Covariate + Interaction					-1.227	-44.087
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	351	351	158	158	509	509
R-squared	0.014	0.407	0.006	0.307	0.014	0.380

Notes: The table shows degree of association between the teacher characteristic and their estimations. Columns (1)-(4) show results from the regression of actual percent-correct scores on the estimated separately for teachers with a pre-service teacher training and for those without. Columns (5) and (6) show the differences between both the groups. The data on teachers' training statuses were collected from the survey of teachers in both India and Bangladesh. Each observation in the estimation is a teacher-student dyad. Standard errors (clustered at the teacher level) are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S12. Predictive power of teacher's experience in India and Bangladesh.

	(1)	(2)	(3)	(4)	(5)	(6)
	Above median overall experience		Below median overall experience		Test for interaction effects	
A. India						
Math						
Estimated	0.333***	0.446***	0.291***	0.451***	0.291***	0.451***
	(0.072)	(0.064)	(0.053)	(0.052)	(0.052)	(0.051)
Covariate					-3.273	12.634**
					(6.213)	(5.637)
Interaction					0.043	-0.004
					(0.088)	(0.082)
Covariate + Interaction					-3.231	12.629
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	219	219	220	220	439	439
R-squared	0.132	0.276	0.131	0.432	0.132	0.348
B. Bangladesh						
Math						
Estimated	-0.000	0.168	0.200***	0.414***	0.200***	0.414***
	(0.094)	(0.210)	(0.075)	(0.139)	(0.075)	(0.139)
Covariate					13.131*	47.406***
					(7.892)	(16.754)
Interaction					-0.200*	-0.246
					(0.120)	(0.251)
Covariate + Interaction					12.931	47.160
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	278	278	327	327	605	605
R-squared	0.000	0.561	0.019	0.579	0.010	0.570
Language						
Estimated	0.096	0.543***	0.148**	0.377***	0.148**	0.377***
	(0.080)	(0.203)	(0.071)	(0.114)	(0.071)	(0.113)
Covariate					1.862	-49.783***
					(6.775)	(13.845)
Interaction					-0.052	0.166
					(0.106)	(0.234)
Covariate + Interaction					1.81	-49.617
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	248	248	261	261	509	509
R-squared	0.005	0.335	0.018	0.426	0.014	0.381

Notes: The table shows degree of association between the teacher characteristic and their estimations. Columns (1)-(4) show results from the regression of actual percent-correct scores on the estimated separately for teachers with an above median overall teaching experience and for those with below median overall teaching experience. Columns (5) and (6) show the differences between both the groups with below median experience as the reference category for the comparison. The data on teachers' experience levels were collected from the survey of teachers in both India and Bangladesh. Each observation in the estimation is a teacher-student dyad. Standard errors (clustered at the teacher level) are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S13. Predictive power of teacher's experience in the school in India and Bangladesh.

	(1)	(2)	(3)	(4)	(5)	(6)
	Above median overall experience in the school		Below median overall experience in the school		Test for interaction effects	
A. India						
Math						
Estimated	0.343***	0.443***	0.237***	0.460***	0.237***	0.460***
	(0.055)	(0.049)	(0.063)	(0.074)	(0.062)	(0.073)
Covariate					-2.255	-9.332*
					(5.759)	(5.499)
Interaction					0.106	-0.018
					(0.082)	(0.087)
Covariate + Interaction					-2.149	-9.349
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	259	259	180	180	439	439
R-squared	0.160	0.327	0.078	0.356	0.146	0.348
B. Bangladesh						
Math						
Estimated	-0.004	0.213	0.217**	0.393***	0.217**	0.393***
	(0.084)	(0.204)	(0.087)	(0.142)	(0.087)	(0.142)
Covariate					12.350	42.999***
					(8.012)	(16.423)
Interaction					-0.221*	-0.180
					(0.121)	(0.247)
Covariate + Interaction					12.128	42.819
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	285	285	320	320	605	605
R-squared	0.000	0.541	0.021	0.591	0.014	0.570
Language						
Estimated	0.064	0.425**	0.183**	0.449***	0.183***	0.449***
	(0.085)	(0.179)	(0.070)	(0.123)	(0.070)	(0.122)
Covariate					6.065	-39.206***
					(6.994)	(12.555)
Interaction					-0.119	-0.024
					(0.110)	(0.217)
Covariate + Interaction					5.945	-39.23
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	271	271	238	238	509	509
R-squared	0.002	0.395	0.029	0.353	0.016	0.379

Notes: The table shows degree of association between the teacher characteristic and their estimations. Columns (1)-(4) show results from the regression of actual percent-correct scores on the estimated separately for teachers with an above median teaching experience in the school that he/she is currently teaching and for those with below median teaching experience in the school. Columns (5) and (6) show the differences between both the groups with below median experience as the reference category forthe comparison. The data on teachers' experience levels were collected from the survey of teachers in both India and Bangladesh. Each observation in the estimation is a teacher-student dyad. Standard errors (clustered at the teacher level) are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S14. Predictive power of teacher's experience teaching the subject in India and Bangladesh.

	(1)	(2)	(3)	(4)	(5)	(6)
	Above median experience teaching the subject		Below median experience teaching the subject		Test for interaction effects	
A. India						
Math						
Estimated	0.356***	0.465***	0.294***	0.400***	0.294***	0.409***
	(0.060)	(0.060)	(0.053)	(0.060)	(0.053)	(0.057)
Covariate					-9.046	-9.581*
					(5.741)	(5.234)
Interaction					0.061	0.087
					(0.077)	(0.076)
Covariate + Interaction					-8.984	-9.494
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	218	218	221	221	439	439
R-squared	0.164	0.442	0.121	0.347	0.148	0.355
B. Bangladesh						
Math						
Estimated	0.056	0.267	0.166*	0.358**	0.166*	0.358**
	(0.086)	(0.184)	(0.087)	(0.155)	(0.087)	(0.155)
Covariate					7.755	60.010***
					(8.077)	(16.466)
Interaction					-0.110	-0.091
					(0.122)	(0.240)
Covariate + Interaction					7.645	59.919
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	299	299	306	306	605	605
R-squared	0.002	0.584	0.013	0.553	0.007	0.569
Language						
Estimated	0.102	0.418**	0.189**	0.455***	0.189**	0.455***
	(0.084)	(0.161)	(0.076)	(0.143)	(0.076)	(0.142)
Covariate					7.847	35.487***
					(7.146)	(11.726)
Interaction					-0.087	-0.037
					(0.113)	(0.215)
Covariate + Interaction					7.76	35.45
Teacher FE	No	Yes	No	Yes	No	Yes
Observations	242	242	267	267	509	509
R-squared	0.008	0.400	0.022	0.358	0.018	0.379

Notes: The table shows degree of association between the teacher characteristic and their estimations. Columns (1)-(4) show results from the regression of actual percent-correct scores on the estimated separately for teachers with an above median experience teaching the given subject (math or language) and for those with below median teaching experience forthe subject. Columns (5) and (6) show the differences between both the groups with below median experience as the reference category for the comparison. The data on teachers' experience levels were collected from the survey of teachers in both India andBangladesh. Each observation in the estimation is a teacher-student dyad. Standard errors (clustered at the teacher level) are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S15. Differences in estimated and actual percent-correct scores of individual students based on student and teacher sex in India.

	(1)	(2)	(3)
A. All teachers	Estimated	Actual	Col. (1)-(2)
Boys	61.46	37.49	23.75***
	[24.76]	[21.40]	(3.09)
Girls	65.52	37.75	27.69***
	[22.37]	[20.82]	(2.86)
N (students)	456	456	912
B. Male teachers			
Boys	64.45	33.28	30.89***
	[19.44]	[17.90]	(4.89)
Girls	67.57	35.26	31.97***
	[17.98]	[17.84]	(8.22)
N (students)	90	90	180
C. Female teachers			
Boys	60.83	38.41	22.22***
	[25.74]	[22.03]	(3.54)
Girls	64.98	38.43	26.55***
	[23.40]	[21.57]	(2.90)
N (students)	366	366	732

Notes: The table presents estimations from the comparison of actual percent-correct scores and estimated percent-correct score separately by student and teacher sex. The estimates are grouped by teachers' sex in different panels. Panel A includes all teachers, panel B only includes male teachers, and panel C only includes female teacher. Column (1) shows teachers' estimations and column (2) shows actual percentages. Column (3) shows the difference between estimated and actual percentages (positive differences indicate over-estimations and negative differences indicate underestimations). Each observationis a student. Standard errors (clustered at the teacher level) are in parentheses and standard deviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S16. Differences in estimated and actual percent-correct scores of individual students based on students' socio-economic status.

	(1) Estimated	(2) Actual	(3) Col. (1)-(2)
Students with...			
Above-median SES	63.59	38.01	25.50***
	[23.04]	[20.69]	(2.55)
Below-median SES	62.49	36.53	25.48***
	[25.11]	[22.35]	(3.41)
N (students)	456	456	912

Notes: The table presents estimations from the comparison of actual percent-correct scores and estimated percent-correct score separately by students' socio-economic status (SES) level. SES level is calculated as the first principal component from the principal component analysis of household asset indications collected from the student survey. Column (1) shows teachers' estimations and column (2) shows actual percentages. Column (3) shows the difference between estimated and actual percentages (positive differences indicate over-estimations and negative differences indicate underestimations). Each observationis a student. Standard errors (clustered at the teacher level) are in parentheses and standard deviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S17. Differences in estimated and actual percent-correct scores of individual students based on students' caste status.

	(1) Estimated	(2) Actual	(3) Col. (1)-(2)
Student from...			
Scheduled castes or tribes	62.19	36.27	25.74***
	[23.60]	[21.20]	(2.62)
Other castes	65.44	38.79	26.51***
	[23.51]	[21.18]	(2.74)
N (students)	456	456	912

Notes: The table presents estimations from the comparison of actual percent-correct scores and estimated percent-correct scoreseparately by students' caste group. Scheduled casts and scheduled tribes are historically disadvantaged social classes in India. Column (1) shows teachers' estimations and column (2) shows actual percentages. Column (3) shows the difference between estimated and actual percentages (positive differences indicate over-estimations and negative differences indicate underestimations). Each observation is a student. Standard errors (clustered at the teacher level) are in parentheses and standard deviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table S18. Differences in estimated and actual percent-correct scores of individual students based on students' intelligence levels and teacher subject knowledge in India.

	(1) Estimated	(2) Actual	(3) Col. (1)-(2)
Students with...			
Above-median intelligence	70.02	42.76	26.96***
	[20.75]	[20.50]	(2.75)
Below-median intelligence	56.10	32.14	23.83***
	[24.60]	[20.73]	(3.12)
N (students)	456	456	912

Notes: The table presents estimations from the comparison of actual percent-correct scores and estimated percent-correct score based on students' fluid intelligence levels as measured by raven's progressive matrices. Column (1) shows teachers' estimationsand column (2) shows actual percentages. Column (3) shows the difference between estimated and actual percentages (positive differences indicate over-estimations and negative differences indicate underestimations). Each observation is a student. Standard errors (clustered at the teacher level) are in parentheses and standard deviations are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.