The Foundational Math Skills of Indian Children*

Andreas de Barros†

Alejandro J. Ganimian‡

Abstract

We leverage data on learning for 101,084 public-school students in grades 4, 6, and 8 across 19 Indian states to diagnose their mathematic skills. These data allow us to diagnose their achievement on less frequently assessed skills. We use a novel approach to estimate the share of students who can meet fourth-grade standards. We find that the foundational skills of children are even lower than previously documented: 52% mastered frequently assessed skills, but only 27% mastered typically unassessed skills. These children also make less progress than believed. Gender gaps in these skills emerge between grades 4 and 6 and persist.

Keywords: foundational skills, gender, India, learning crisis, mathematics

JEL: I21 – Analysis of Education; I25 – Education and Economic Development; O15 – Human Resources; Human Development; Income Distribution; Migration

---

* We gratefully acknowledge funding from the Douglas B. Marshall, Jr. Family Foundation for this project. We especially thank Karthik Muralidharan for his initial involvement in this project as a collaborator. We thank Felipe Barrera-Osorio, Masha Bertling, Emmerich Davies, Andrew Ho, Jalnidh Kaur, Edward Kim, Pranav Kothari, Aarthi Muralidharan, Sridhar Rajagopalan, Maulik Shah, Nishchal Shukla, Gayatri Vaidya, and Martin West for useful feedback that informed this study. We thank Educational Initiatives for making data available. Google.org supported Educational Initiatives with funding for the Student Learning Study. We also thank Anuja Venkatachalam for excellent research assistance. The usual disclaimers apply. The authors have no conflicting interests to declare.

† Postdoctoral Associate at the Massachusetts Institute of Technology's Department of Economics. E-mail: debarros@mit.edu.

‡ Assistant Professor of Applied Psychology and Economics at New York University's Steinhardt School of Culture, Education, and Human Development. E-mail: alejandro.ganimian@nyu.edu.

1. Introduction

Over the past 17 years, the non-profit Pratham has conducted annual assessments of children's language and math skills across rural India (Banerji, 2015). These Annual Status of Education Reports (ASER) have played a pivotal role in shaping domestic and global education-policy debates, given that one in five school-age children lives in India (World Bank, 2019).

In math, the ASER assessments—which focus on number recognition, subtraction, and division—have documented four very important patterns. First, many children start primary school without recognizing numbers: 36% of children in grade 1 cannot recognize single-digit numbers (ASER, 2018). Second, most children graduate from primary school without performing basic arithmetic: 48% of students in fifth grade (i.e., the last year of primary school) cannot solve two subtractions of a two-digit number by another or a division of a three-digit number by a one-digit number (ASER, 2018). Third, these two facts have changed relatively little over the past decade: in 2005, the share of first graders who could not recognize numbers was 58% and the share of fifth graders who could not solve arithmetic operations was 27% (ASER, 2006). Fourth, math skills vary widely across India. For example, while nearly 51% of fifth-graders in Haryana can do division, less than 18% of those in Assam can (ASER, 2018).

The focus of these assessments on number recognition and arithmetic has had several advantages. First, it rendered the assessments relatively simple to administer, enabling universities, teacher-training institutions, and other community organizations to do so, not only evaluating over a million children annually, but also building local assessment capacity (ASER, 2014). Second, it raised public awareness of the abysmally low learning levels in the country, generating consensus about what is now widely acknowledged as a "learning crisis" (Pritchett, 2013). Lastly, it made the ASER tests useful data-collection instruments for impact evaluations of education interventions, which have been key to identifying effective programs (e.g., Banerjee, Banerji, Duflo, Glennerster, & Khemani, 2010; Banerjee, Banerji, Duflo, & Walton, 2011; Banerjee, Cole, Duflo, & Linden, 2007; Duflo, Berry, Mukerji, & Shotland, 2015).[1]

Yet, the focus on number recognition and arithmetic has arguably also detracted attention from other foundational skills, such as fractions and decimals, geometry, and measurement. The

[1] In fact, it is in great part due to these advantages that the ASER tools are now used across multiple developing countries in South and East Asia, Sub-Saharan Africa, and Latin America. For an up-to-date list of the People's Action for Learning (PAL) network, visit: https://palnetwork.org.

only other periodic representative assessment of children's math skills in India, the National Achievement Survey (NAS)—a survey of a sample of roughly two million students in grades 3, 5, 8 and 10 conducted by the National Council of Education Research and Training (NCERT)—includes domains such as measurement and geometry. Yet, the fact that the NAS results are consistently more sanguine than ASER and other domestic studies has raised questions about whether they offer accurate and reliable estimates of children's learning (see CABE, n.d., p. 58). Some cross-sectional and longitudinal studies have offered richer descriptions of the math skills of Indian children (see, e.g., Banerjee et al., 2022; Das & Zajonc, 2010; Muralidharan, Singh, & Ganimian, 2019; Muralidharan & Sundararaman, 2011; Singh, 2019), their focus on specific cities and states does not allow for a national diagnostic of children's foundational math skills.

Evidence from developed countries suggests that fractions and decimals, geometry, and measurement matter for children's learning during school and beyond. Students who struggle with fractions have trouble making progress in math and related areas, and are likely to face difficulties as adults. It is crucial that they are not only exposed to basic geometric shapes, names, and concepts early in their schooling, but that they transition from concrete to abstract representations (National Mathematics Advisory Panel, 2008). The importance of these skills was recently recognized in India's new National Education Policy (MHRD, 2019).

In this paper, we try to address this gap by presenting detailed, representative, and previously unpublished, learning outcomes data on 101,084 public-school students across 18 Indian states and one union territory. According to the 2011 Indian Census, the area covered by our data represents 861.2 million individuals (MHA, 2012)—or 12% of the world's population. These data were collected as part of the Student Learning Survey (SLS), conducted by Educational Initiatives, a leading assessment firm in India, in collaboration with state governments in 2009. They cover three grades spanning elementary and middle school (grades 4, 6, and 8) and include not only the arithmetic skills reported by ASER (number recognition and arithmetic), but also three other skills on which—to our knowledge—there has been no prior reporting at the national level in India (fractions and decimals, measurement, and geometry).[2]

[2] We decided to focus on math because, while similar data are available for language, it is more challenging to compare language skills across multiple states in which students speak a variety of vernacular languages.

We take advantage of the fact that students across grades are assessed on the same skills to map their performance onto a common scale. The analytical approach we use, known as Cognitive Diagnostic Models (CDMs), allows us to compare the performance of fourth, sixth, and eighth graders side by side. Specifically, we express the performance of all students in terms of whether they have mastered the skills expected of a typical fourth grader.[3] This comparison cannot be achieved by calculating the proportion of items assessing each skill that were answered correctly by students at each grade because each test contains a very small number of common items for each skill.[4] It cannot be achieved by using other analytical methods based on Item Response Theory (IRT) because there are few items overall for each skill.[5]

We present three main findings. First, while we confirm that primary- and middle-school students perform poorly in the basic skills regularly assessed by ASER, they fare even worse on other foundational skills not captured by those tests: 52% of all students in our data have mastered number sense and arithmetic, but only 27% have mastered fractions and decimals, measurement, and geometry. Second, while student achievement in previously assessed skills improves across grades, the corresponding trajectory in the previously unassessed skills is flat: the percentage of students who have mastered number sense and arithmetic increases—from 43% in grade 4, to 50% in grade 6, to 61% in grade 8—whereas the share of students who have mastered the remaining three skills remains virtually unchanged—from 22% in grade 4, to 28% in grade 6, and 29% in grade 8. Third, girls perform below boys for these previously unassessed skills. The gap emerges between grades 4 and 6 and remains unchanged in grade 8. In grade 4, the percentage of boys and girls who have mastered these skills is equal (22%); by grade 6, 25% of girls and 31% of boys have done so, and this gap persists in grade 8.

This study makes three main contributions to existing evidence on the foundational math skills of Indian children. First, it offers one of the most geographically representative accounts to date of the performance of such children that goes beyond number recognition and arithmetic

[3] We do so for ease of interpretation. Our results can be interpreted as indicating whether students in primary and secondary schools have mastered basic skills.

[4] Across the fourth-grade, sixth-grade, and eighth-grade assessments, there are only two common items that were administered to all students. For example, comparisons for the "fractions and decimals" skill would therefore rest on a single test question.

[5] For example, the fourth-grade assessment includes just five questions mapped to the "fractions and decimals" skill.

operations, showing that the learning crisis in India may go farther than previously shown—affecting fractions and decimals, geometry, and measurement more severely. Second, it builds on prior studies that track students' achievement through the school system (see Muralidharan et al., 2019; Muralidharan & Zieleniak, 2014; Pritchett & Beatty, 2015), documenting that growth in these less frequently assessed skills may be even slower than in those assessed more regularly. Finally, to our knowledge, it is one of the few studies that disaggregate achievement in these skills by students' sex, raising the urgency of addressing gender inequality in the country—particularly, given the importance of science, technology, engineering, and math skills in the Indian economy (World Bank, 2018).

2. Sample

The data that we use in this study were collected as part of the Student Learning Survey (SLS), which was conducted by Educational Initiatives (EI), a leading assessment firm in India, in collaboration with state governments between January and September 2009. EI recruited 18 major Indian states and one union territory for this study due to their population size: they each counted with more than one percent of India's total population (of 1.03 billion, as per India's 2001 census). Figure 1 shows the participating locations. The study focused on public schools because most Indian students attend public schools.[6] It assessed grades 4, 6, and 8 to measure learning at different stages of students' schooling trajectory, including: lower-primary school (grade 4), upper-primary school (grade 6), and middle school (grade 8).[7] The sampling frame for the study included 421 districts and their 657,787 government-run schools, with a collective enrolment of 25,519,296 students across these three grade levels.[8]

The sample for the study was representative of the student population in the participating states. EI drew a two-step, random stratified cluster sample as follows. First, within each state, it categorized districts by level of development, and it randomly selected two to four districts

---

[6] As of the 2015-16 school year, 74% of primary and upper-primary schools in India are public schools (Mehta, 2017) and 65% of primary-school (i.e., both lower- and upper-primary school) students were served by public schools (UNESCO Institute for Statistics 2018).

[7] Grade 8 also marks the end of free education, as per India's Right to Education Act (RTE).

[8] The sample in this study represents roughly 72% of the total Indian government school population in these grade-levels.

across those levels (depending on the size of the state), for a total of 48 districts. Then, it randomly selected 2,399 schools across those districts, through a process in which schools with more students were more likely to be selected (this process is known as "probability proportional to size" or PPS sampling).[9] All students in grades 4, 6, and 8 who were present on the day of the survey were invited to participate. The total sample included 101,084 students: 29,513 students in grade 4, 35,604 in grade 6, and 35,967 in grade 8. Approximately 67% of enrolled students took the math test (EI, 2010). This percentage may seem low, but it is actually similar to the share of students attending school regularly, which matches similar net attendance and absenteeism rates reported elsewhere (see ASER, 2018; IIPS, 2007).[10] We provide additional details on the SLS in Appendix A, including a comparison with other large-scale assessments from India (Singh, 2020).

3. Data

The dataset for this study includes students' responses to each item of the math assessments administered in grades 4, 6, and 8. EI designed these assessments as follows. First, it reviewed the curricula and official textbooks used in participating states and union territories to understand what students ought to know and be able to do by the end of each primary-school grade and identify common expectations across states. Second, it convened workshops with subject-specific and assessment experts to finalize the competencies to be assessed and develop items. Third, it created three versions of each subject test, to prevent student cheating, and translated it into three languages (Hindi, Telugu, and Gujarati) for pilot testing. Fourth, it piloted the assessments in three districts (Ghaziabad, Uttar Pradesh; Medak, Andhra Pradesh; and Vadodara, Gujarat) and conducted interviews with teachers to analyze the test and question (hereafter, "item") characteristics and make adjustments to ensure they were appropriate for the broader sample. This pilot included 24,600 students across 30 towns in the three districts. Lastly,

[9] For more information on the sampling procedure, see EI (2010). One of the main advantages of PPS sampling is that it is "self-weighting"—i.e., each sub-unit has equal probability of selection, so that no reweighting is necessary for estimation purposes (see Skinner, 2006).

[10] A narrative description of the assessment suggests student non-response was almost exclusively due to student absenteeism, and not due to students' refusal to participate in the test (EI, 2010). Unfortunately, however, we do not have detailed data on non-response reasons.

the final versions of the tests were translated into the 13 languages of the target sample seeking to preserve the original meaning of the question, the reading level of the text, and the difficulty level of each item. The translated versions were then reviewed by language experts and back-translated and, after several iterations, the tests were back-translated into English. All tests were administered to entire classrooms in a written format in blocks of 120 minutes per subject (for a detailed description of design and administration and test papers, see EI, 2010).

A distinguishing feature of this dataset is that EI mapped each item to one or more of five content domains in math: (a) number concepts and theory (e.g., completing a missing number in a sequence of two-digit numbers); (b) operations on whole numbers (e.g., subtracting a two-digit number from another); (c) fractions and decimals (e.g., identifying the fraction represented a shaded part of a figure); (d) measurement (e.g., measuring the length of a pencil with a ruler); and (e) shapes and geometry (e.g., distinguishing a triangle from other shapes).[11] As we discuss in the next section, we are interested in expressing the results of these assessments in terms of what a fourth grader is expected to know and is able to do, so we drop items for domains that are not taught in grade 4 (e.g., algebra). After discarding these items, we end up with 86 unique items: 40 of them were administered in grade 4, 41 in grade 6, and 28 in grade 8. Importantly, 20 of these are common across any two grades. These "anchor" items (i.e., common items across tests) allow us to map the performance of all students onto a common scale (see Table C.1 in Appendix C).

Another important feature of this dataset is that EI also mapped each item to a grade level (based on the curriculum and textbook reviews).[12] For example, an item may assess whole-number operations at a fourth-grade level (e.g., 76+27), at a sixth-grade level (e.g., 713x24), or at an eighth-grade level (e.g., (-6x-5)-6+5). This level of specificity is crucial for our analytical strategy because we leverage it to express the performance of all students with respect to grade 4 curricular standards for math (see Table C.1 in Appendix C for the number of items by grade and content domain).

[11] Note that there is no clear ordering or developmental progression of these skills, in terms of lower- vs higher-level abilities. For example, there is an active area of research on the relationship between geometry and spatial thinking on the one hand and operations on whole numbers on the other (Frye et al., 2013; Hodgen et al., 2020).

[12] We follow Educational Initiatives' mapping of test questions to grade levels.

4. Analytical strategy

Our analytical strategy allows us to estimate whether each student has "mastered" (or is "proficient" in) each of the five mathematical skills mentioned above, at a fourth-grade level. Other commonly used approaches, such as "classical test" or "item response" theory, seek to estimate each student's proficiency on a single (e.g., math) domain, as a function of that student's "latent" (i.e., unobserved) ability and the characteristics of the items on a test (Andersen, 1983).[13] The approach we use, known as a "cognitive diagnostic model" (CDM), seeks to estimate each student's proficiency for a set of related but separable (e.g., numbers, operations, measurement) domains (de la Torre & Chiu, 2016).[14] The two main advantages of CDMs are their potential to integrate theories of cognition in the scoring of students' performance on a test and their capacity to make judgments about individual students' performance without regard to their relative standing with respect to other examinees (see de la Torre et al., 2016).

Overall, our approach entails four main steps. First, we map each item to a set of skills being assessed; each item can be mapped to multiple skills.[15] In this study, we have obtained this mapping (known in this literature as a "Q-matrix") directly from the test developers. Then, we specify a model of how these skills may determine a student's probability of answering an item correctly. In our model, mastering a given skill may affect this probability independently of other skills (this would be considered a "main effect"), the skill may also affect the probability in conjunction with the student's knowledge of other skills (this would be an "interaction effect"), or students may simply guess the correct answer (these three are known as "item parameters"). Next, we estimate the model's parameters with our data. This estimation is iterative: In one turn, it tries to improve its estimates of the aforementioned item parameters; in another turn, it

---

[13] We cannot use an Item Response Theory-based approach for our purposes, because of the low number of test questions per domain.

[14] Alternatively, while classical test theory models seek to model test scores, item response theory models aim to model test *items*, and cognitive diagnosis models try to model the components of reasoning required to answer specific items (de la Torre, Carmona, Kieftenbeld, Tjoe, & Lima, 2016). CDMs are also known as "diagnostic classification models" or DCMs.

[15] CDMs do not require an equal number of test questions across skills. A larger number of test questions per skill allow for greater precision in the estimation of student ability; yet, a lower number of questions does not introduce bias in the estimation of student skill profiles.

calculates the expected count of students who fall into any given “skill class” (i.e., the possible combinations of all skills assessed in the test). Finally, armed with all the information above, we categorize each individual student into one of these skill classes.[16]

This estimation seeks to determine the skill class to which each student belongs. We focus on skill classes at the fourth-grade level. However, to account for the fact that some items cover materials beyond grade 4, we introduce five additional, ancillary categories (one for each of the five skills). They reflect that a student may be proficient in material beyond grade 4.[17] With these three possible mastery levels (mastery in material beyond grade 4, mastery of material at a fourth-grade level, and below) and five math skills covered by the test, each student can be categorized into one of 3^5 or 243 possible skill classes. 2^5 or 32 of these classes indicate mastery at a fourth-grade level or beyond—our main variable of interest. We are thus able to express the performance of students across all three grades in our study (i.e., grades 4, 6, and 8) with respect to mastery of fourth-grade curricular expectations (e.g., the percentage of students who have mastered fourth-grade arithmetic).

Formally, for each item i on the test, we let the vector $\boldsymbol{q}_i = (q_{i1}, q_{i2}, q_{i3}, \dots, q_{iK})$ represent whether the item measures ($q_{ik} = 1$) or does not measure ($q_{ik} = 0$) each math skill, such that with I items and K skills, we can construct an $I \times K$ Q-matrix mapping items to skills. Further, for each student e, we let the vector $\boldsymbol{\alpha}_e = (\alpha_{e1}, \alpha_{e2}, \alpha_{e3}, \dots, \alpha_{eK})$ represent the student’s mastery ($\alpha_{ek} = 1$) or non-mastery ($\alpha_{ek} = 0$) of each math skill $k = 1, \dots, K$ assessed in the test. The vectors $\boldsymbol{q}_i$ and $\boldsymbol{\alpha}_e$ are similar, but the item vectors are considered to be known whereas the student vectors are unobserved (and must thus be estimated).[18] In our case, $K = 10$ (five categories of interest and 5 ancillary categories).[19]

---

[16] Note that our explanation here is simplified. For additional details, see Rupp, Templin, and Henson (2010).

[17] We also investigated scenarios that allow for students to master higher-grade material and forget material from earlier grades. We did not find this phenomenon to be prevalent. We also prefer our approach because of its lower number of possible classes (243 vs. 1024), which leads to a lower probability of mis-classifying a student.

[18] In this study, this matrix was composed by EI (i.e., the test developers) based on theoretical work, qualitative research, and their subject-matter experts. Yet, there are multiple approaches to compose and validate Q-matrices (see de la Torre & Chiu, 2016).

[19] We decided to use odd entries in a vector to refer to grade-four skills, and even entries in a vector to refer to higher-grade skills. For example, the vector $(1,1,0,0,0,0,0,0,0,0)$ may refer to a student who has mastered fractions (both at a grade-four level and beyond), but none of the remaining skills on the test (not even at a grade-four level).

With this setup, we estimate a student's probability of solving a given fourth-grade item by fitting the following linear probability model:

$$P(X_i = 1|\alpha_{ei}^*) = \lambda_i + \lambda_{i1}\alpha_{e1} + \lambda_{i2}\alpha_{e2} + \lambda_{i(1*2)}\alpha_{e1}\alpha_{e2}, \quad (1)$$

where λ_i indicates a student's probability of solving the item correctly; λ_{i1} reflects a student's increase in probability if they have mastered the first skill mapped to the item; λ_{i2} represents the increase in that probability if the student has mastered the second skill mapped to the item; and $\lambda_{i(1*2)}$ indicates the increase in that probability due to potential complementarities across the two skills. In this model, we only include one interaction term because all fourth-grade items are mapped to a maximum of two skills.

In turn, we model a student's probability of solving a given higher-grade item by fitting:

$$P(X_i = 1|\boldsymbol{\alpha}_{ei}^*) = \lambda_i + \lambda_{i1}\alpha_{e1} + \lambda_{i2}\alpha_{e2}, \quad (2)$$

where λ_{i1} reflects a student's increase in probability if they have mastered a fourth-grade understanding of the skill mapped to the item; and λ_{i2} reflects an increase in probability if they have mastered a higher-grade understanding of the skill mapped to the item. Everything else is as in equation (1). This model does not include an interaction term because all items capturing higher-level materials are mapped to a single skill, at its two levels. We discuss additional technical details in Appendix B.

Once we obtain each student's skill profile, we calculate the proportion of students who have mastered each of the five skills mentioned above at a fourth-grade level. We compare the percentage of students who have achieved this level of mastery on previously assessed skills (i.e., number concepts and theory and operations on whole numbers) to the percentage who reached it on previously unassessed skills (i.e., fractions and decimals, measurement, and shapes and geometry) to determine whether students' performance on the former is higher than on the

---

We do not allow for even entries to be 1 if odd entries are 0 (see footnote 19). For a discussion of polytomous CDMs, see de la Torre et al. (2016).

latter. We also compare the percentage of students at this level of mastery across male and female students.

5. Results

We present three sets of results. First, we describe the extent to which students in our sample achieve a fourth-grade proficiency level in the five skills assessed by the math test. Specifically, we show that students perform better in previously assessed skills than in the previously unassessed skills. Then, we describe how students' performance on those skills varies by grade. Students improve faster in previously assessed skills than in previously unassessed skills. Finally, we present how students' performance varies by sex. Gaps by student sex emerge at the end of primary school and persist in middle school.

Performance on previously assessed and unassessed skills

Our analysis indicates that primary- and middle-school students perform poorly in skills typically assessed by other tests, but they fare even worse in less commonly assessed skills (Figure 2). Specifically, whereas 69% of students in grades 4, 6, and 8 achieve fourth-grade proficiency in number sense and 60% in operations (two skills frequently assessed by other assessments) the corresponding mastery rates for fractions, geometry, and measurement (three skills less commonly assessed) are 55%, 56%, and 50%, respectively. In fact, whereas 52% of students across these three grades achieve fourth-grade proficiency in both previously assessed skills, only 27% of them reach such mastery on the three less commonly assessed skills. This contrast is partly due to the number of skills included in each category (i.e., reaching mastery in three skills is, by definition, more difficult than doing so in two skills). Yet, the difference is not entirely driven by grouping, as we demonstrate in Figure C.1 in Appendix C. Thus, focusing on number sense and operations may convey a more optimistic picture of what Indian students know and are able to do than a more comprehensive assessment.

Performance by students' grade

Students in middle school are more likely to master commonly assessed skills than those in primary school, but they are only slightly more likely to master previously unassessed skills (Figure 3). In other words, the percentage of students achieving fourth-grade mastery of number sense and operations (two skills frequently assessed) increases from 43% in grade 4 to 50% in grade 6 to 61% in grade 8. However, the percentage of students reaching this proficiency level in fractions, geometry, and measurement (three skills less frequently assessed) increases much more slowly from 22% in grade 4 to 28% in grade 6 to 29% in grade 8. By focusing on number sense and operations, prior diagnostics may have overestimated the progress that Indian students make between primary and middle school.[20]

Performance by student sex

Finally, achievement gaps by student sex are more pronounced in previously assessed skills than in previously unassessed skills. Specifically, boys perform 6.6 percentage points (pp.) better than girls in the former, but only 4.3 pp. in the latter (Figure 4). However, these aggregates mask differences in gaps across specific skills. For example, among previously assessed skills, the gap for number sense is much larger (6.8 pp.) than the one for operations. Similarly, among previously unassessed skills, the gap for measurement (7.5 pp.) is twice or more than the respective gap for fractions (3.3 pp.) and geometry (2.6 pp.)

Yet, achievement gaps by student sex widen by different magnitudes as students transition from primary to middle school (Figure 5). In number sense and operations, the gap is already wide in grade 4 (5 pp.) and it widens only slightly in grades 6 (7 pp.) and 8 (8 pp.) In fractions, geometry, and measurement, the gap starts small in grade 4 (less than 1 pp.), but it widens by grade 6 (to 6 pp.), and it remains at this level by grade 8. Achievement gaps in previously unassessed skills emerge later than in more frequently assessed skills.[21]

6. Conclusion

[20] In Appendix Figure C.2, we show the same grade-wise comparison for each of the underlying skills. Students make the greatest progress in operations (a previously assessed skill). The percentage of students reaching proficiency in this skill increases from 49% in grade 4 to 59% in grade 6 to 71% in grade 8.

[21] Recall that very few grade 4 students master those previously unassessed skills; therefore, the lack of an achievement gap may appear unsurprising. At the same time, ex-ante, there is no clear reason to expect the absence of an achievement gap in grade 4 for the previously unassessed skills (which we do observe for the previously assessed skills). Also, the absence of an achievement gap at grade 4 and the onset of learning in the later grades does not make an achievement gap inevitable. If both male and female students improved on these skills at the same rate, then the gap would still be close to zero.

In this paper, we present new evidence on math skills that are not frequently assessed for a representative sample of students in primary and middle schools in India. We capitalize on a large-scale assessment conducted by one of the country's leading assessment firms in collaboration with state and union governments. We employ an innovative analytical approach to understand how learning outcomes evolve along the schooling trajectory. We document three important and novel findings: first, primary- and middle-school students perform even more poorly in less frequently assessed skills (e.g., fractions, geometry, and measurement) than on more frequently assessed skills (e.g., number sense and operations); second, students make less progress in less frequently assessed skills as they move from primary to middle school than in more frequently assessed skills; and finally, girls are at a greater disadvantage vis-à-vis boys in more frequently assessed skills, but achievement gaps by student sex emerge later in children's schooling and remains unchanged by middle school.

Our paper makes an important contribution to mounting evidence on the achievement of Indian children and youth (e.g., ASER, 2016; ASER, 2018; Bhattacharjea, Wadhwa, & Banerji, 2011). Our findings suggest that, while skills such as number sense and whole-number operations may indeed be foundational and easier to measure at scale, an exclusive focus on these skills may present an incomplete picture of what Indian children know and are able to do in math. Such a focus would underestimate the extent of the learning crisis in the Indian education system by failing to acknowledge the poor performance of Indian students on skills such as fractions, geometry, and measurement, overstating the progress that such students make across their schooling trajectory, and incorrectly specify the magnitude of achievement gaps by student sex in math.[22] We believe it would be useful to conduct similar analyses for students' language skills.

Our paper also contributes to the scarce but growing global evidence on learning profiles (e.g., Muralidharan et al., 2019; Muralidharan & Zieleniak, 2014; Pritchett & Beatty, 2015). To our knowledge, ours is the first paper to document how learning outcomes on fractions, geometry, and measurement evolve across the schooling trajectory in a representative sample of

[22] To be clear, we are not arguing that these skills have not been assessed at all before, nor that experts in assessment in India do not understand the importance of student mastery of these content areas. We simply contend that previous analyses of learning outcomes in India have focused on number sense and whole-number operations and that they would do well to expand this focus to less frequently assessed skills.

Indian children and youth. In fact, we believe it may be one of the largest studies addressing this question. The main limitation of this analysis, however, is that we did not track a single cohort of students over time, but rather assessed a cross-section of students across grades. We would welcome analyses that are able to verify the patterns we document in this study with longitudinal data across primary- and middle-school students.[23]

We conclude with three implications for educational policy. First, the *depth* of the global learning crisis may have been severely underestimated: An even larger percentage of Indian students have not mastered foundational math skills. The results therefore lend even greater urgency to a policy shift from student enrolment to student learning. Second, the *scope* of the crisis may be wider than previously known: Students' inability to master foundational skills goes beyond those subskills measured by previous analyses. Our findings support recent policy efforts that place greater emphasis on children's development of foundational skills—but they also imply that such policies may be misdirected if they solely focus on a narrowly defined subset of skills (that only captures number sense and basic arithmetic). Finally, as policymakers shift their focus to learning outcomes and develop interventions to foster children's foundational skills, they look for tangible measures that allow them to track progress. Our article highlights the limitations of commonly used tests for this purpose and provides an example of how assessments can be leveraged to obtain fine-grained indicators of student proficiency levels.

[23] Such studies may also explore potential selection effects if students move from government schools to private schools. Muralidharan and Sundararaman (2015) suggest better performing students select into private schools. However, Kingdon (2020) suggests the overall percentage of India's private school enrollment remains roughly comparable across ages 6 and 18, with no major differences between primary and secondary school.

References

Andersen, E. B. (1983). Latent trait models. *Journal of Econometrics, 22*(1-2), 215-227.

ASER. (2006). *Annual status of education report (ASER) 2005: Final report*. ASER Centre. New Delhi, India.

ASER. (2014). *Instruction booklet*. ASER Centre. Delhi, India.

ASER. (2016). *Middle schools in India: Access and quality. Endline report*. ASER Centre. Delhi, India.

ASER. (2018). *Annual status of education report (ASER) 2018: Provisional*. ASER Centre. New Delhi, India.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy, 2*, 1-30. doi:10.1257/pol.2.1.1

Banerjee, A. V., Banerji, R., Duflo, E., & Walton, M. (2011). *What helps children to learn? Evaluation of Pratham's Read India program in Bihar & Uttarakhand*. Retrieved from Cambridge, MA:

Banerjee, A. V., Bhattacharjee, S., Chattopadhyay, R., Ganimian, A. J., Duflo, E., & Spelke, E. S. (2022). *Street smart or school smart? The arithmetic skills of working children in two Indian cities*. Unpublished manuscript. New York University. New York, NY.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics, 122*, 1235-1264. doi:10.1162/qjec.122.3.1235

Banerji, R. (2015). Building a movement, assessment to action. *Ideas for India*.

Bhattacharjea, S., Wadhwa, W., & Banerji, R. (2011). *Inside primary schools: A study of teaching and learning in rural India*. ASER. Delhi, India.

CABE. (n.d.). *Report of Central Advisory Board of Education sub-committee for assessment and implementation of Continuous and Comprehensive Evaluation (CCE) in the context of the no-detention provision in the Right of Children to Free and Compulsory Education Act, 2009*. Ministry of Human Resource Development, Government of India. New Delhi, India.

Das, J., & Zajonc, T. (2010). India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics, 92*(2), 175-187.

de la Torre, J., Carmona, G., Kieftenbeld, V., Tjoe, H., & Lima, C. (2016). Diagnostic classification models and mathematics education research: Opportunities and challenges.

In A. Izsák, J. T. Remillard, & J. Templin (Eds.), *Psychometric methods in mathematics education: Opportunities, challenges, and interdisciplinary collaborations*. Reston, VA: National Council of Teachers of Mathematics.

de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika, 81*(2), 253-273.

Duflo, E., Berry, J., Mukerji, S., & Shotland, M. (2015). *A wide angle view of learning: Evaluation of the CCE and LEP programmes in Haryana, India*. (Impact Evaluation Report No. 22). International Initiative for Impact Evaluation (3ie). New Delhi, India.

EI. (2010). *Student learning study: Status of student learning across 18 states of India in urban and rural schools*. Ahmedabad, Gujarat: Educational Initiatives (EI).

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide (NCEE 2014-4005)*. National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from http://whatworks.ed.gov

Hodgen, J., Barclay, N., Foster, C., Gilmore, C., Marks, R., & Simms, V. (2020). *Early years and key stage 1 mathematics teaching: Evidence review*. Education Endowment Foundation (EEF). London, UK.

IIPS. (2007). *India National Family Health Survey (NFHS-3), 2005-06*. International Institute for Population Sciences (IIPS). Mumbai, Maharashtra. Retrieved from https://dhsprogram.com/pubs/pdf/frind3/frind3-vol1andvol2.pdf

Kingdon, G. (2020). The private schooling phenomenon in India: A review. *Journal of Development Studies, 56*(10), 1-23.

Mehta, A. C. (2017). *Elementary Education in India. Analytical Tables 2015-16*. Analytical Tables. District Information System for Education (DISE). New Delhi, India. Retrieved from http://udise.in/Downloads/Publications/Documents/Analytical_Table_2015-16.pdf

MHA. (2012). *15th Census of India*. URL: http://www.censusindia.gov.in/ (last accessed: February 15, 2016). Office of the Registrar General & Census Commissioner, Ministry of Home Affairs, Government of India. New Delhi, Delhi.

MHRD. (2019). *National education policy 2020*. Ministry of Human Resource Development, Government of India. New Delhi, India.

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review, 109*(4), 1426-1460.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *The journal of political economy, 119*, 39-77. doi:10.1086/659655

Muralidharan, K., & Sundararaman, V. (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in India. *The Quarterly Journal of Economics, 130*(3), 1011-1066. doi:10.1093/qje/qjv013

Muralidharan, K., & Zieleniak, Y. (2014). *Chasing the syllabus: Measuring learning trajectories in developing countries with longitudinal data and item response theory*. Unpublished manuscript. University of California, San Diego. San Diego, CA.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Pritchett, L. (2013). *The rebirth of education: From 19th-century schooling to 21st-century learning*. Washington, DC: Brookings Institution Press for Center for Global Development.

Pritchett, L., & Beatty, A. (2015). Slow down, you're going too fast: Matching curricula to student skill levels. *International Journal of Educational Development, 40*, 276-288.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications. Methodology in the social sciences*. New York, NY: Guilford Press.

Singh, A. (2019). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, 1-44.

Singh, A. (2020). *Myths of official measurement: Auditing and improving administrative data in developing countries*. (RISE Working Paper No. No. 20/042). Research on Improving Systems of Education (RISE).

Skinner, C. J. (2006). Probability proportional to size (PPS) sampling. In *Encyclopedia of statistical sciences* (pp. 1-5): Wiley Online Library.

World Bank. (2018). *World development report 2018: Learning to realize education's promise*. Washington, DC: The World Bank.

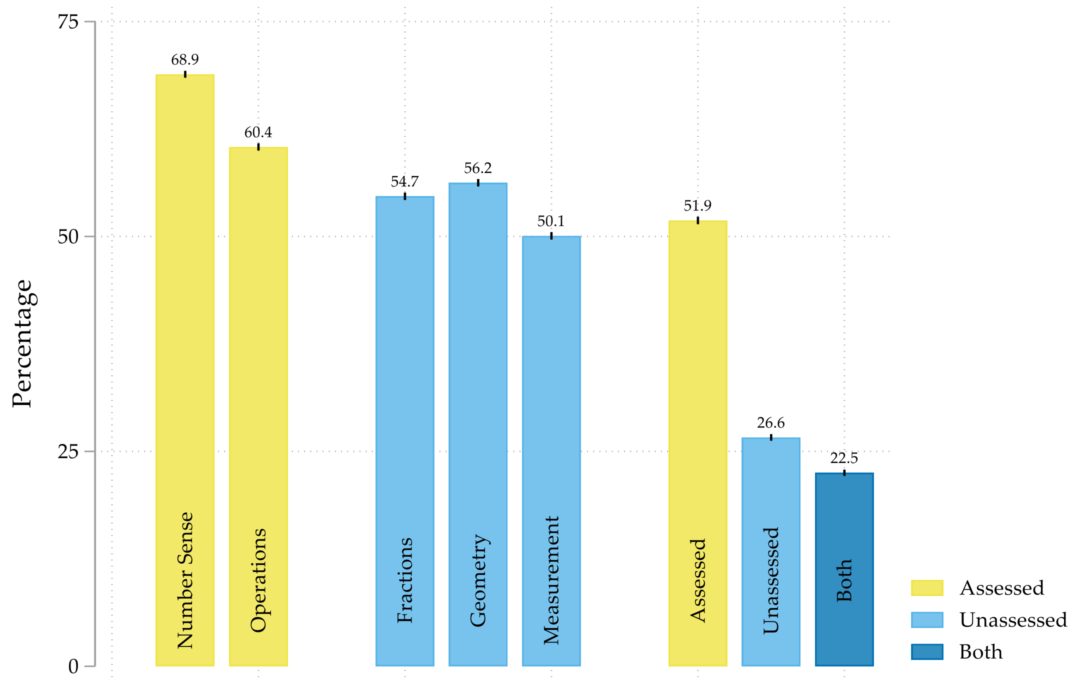
World Bank. (2019). World Bank open data. Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/889464-wbopendata-stata-module-to-access-world-bank-data

Figures

Figure 1: Map of Indian states and union territories participating in the study

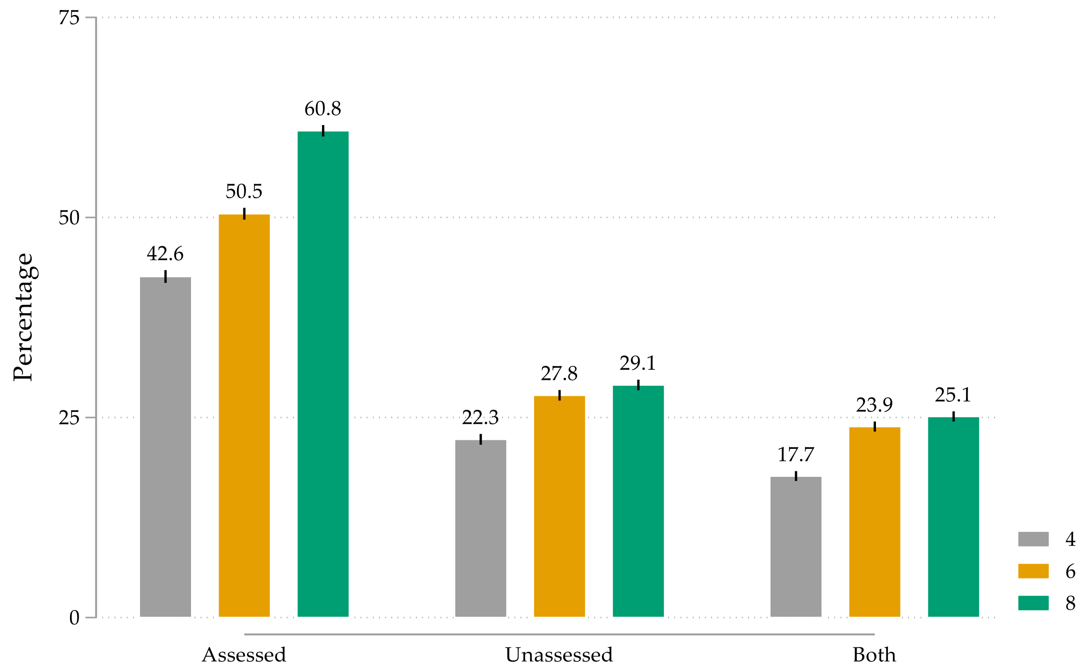
Notes: This map depicts the Indian states and union territories that participated in the study. Dark grey shading highlights participating locations; light grey shading represents the opposite.
Source: Authors' elaboration.

Figure 2: Percentage of students who are proficient in previously assessed and unassessed skills



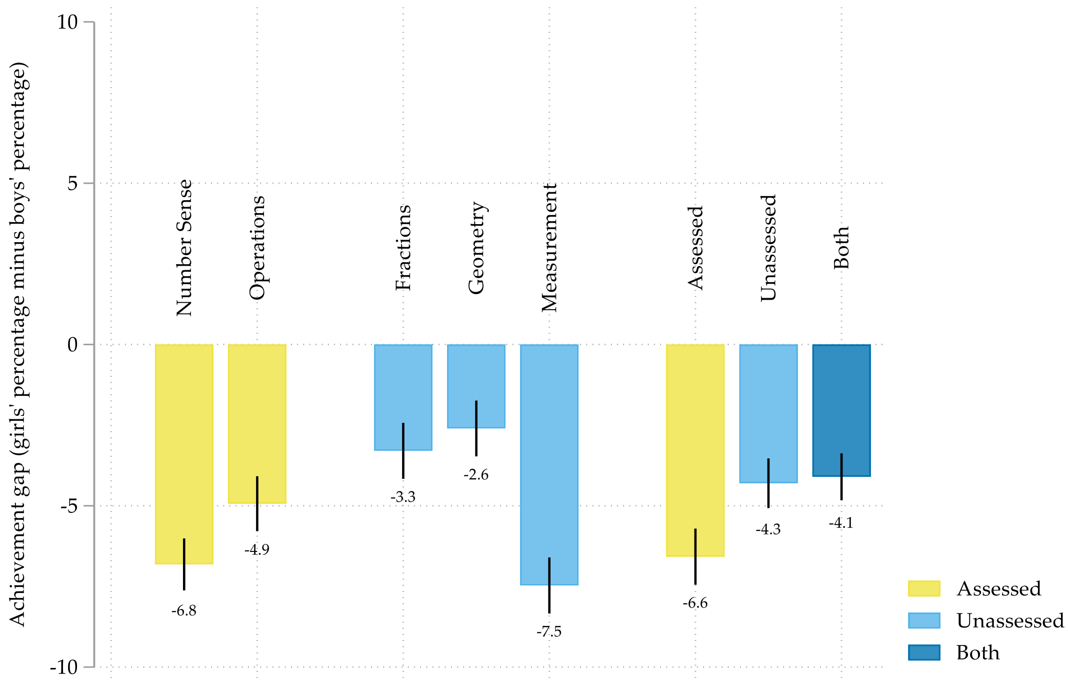
Notes: This figure indicates the percentage of students who have mastered the skills on the test, at a fourth-grade level. "Assessed" refers to skills previously assessed by other assessments: number sense and operations. "Unassessed" refers to skills not previously assessed by other assessments: fractions, geometry, and measurement. "Both" refers to both types of skills. The three bars to the right report on joint mastery of the skills that fall into these three categories. Black vertical bars show 95% confidence intervals.
Source: Authors' elaboration.

Figure 3: Percentage of students who are proficient in previously assessed and unassessed skills, by grade



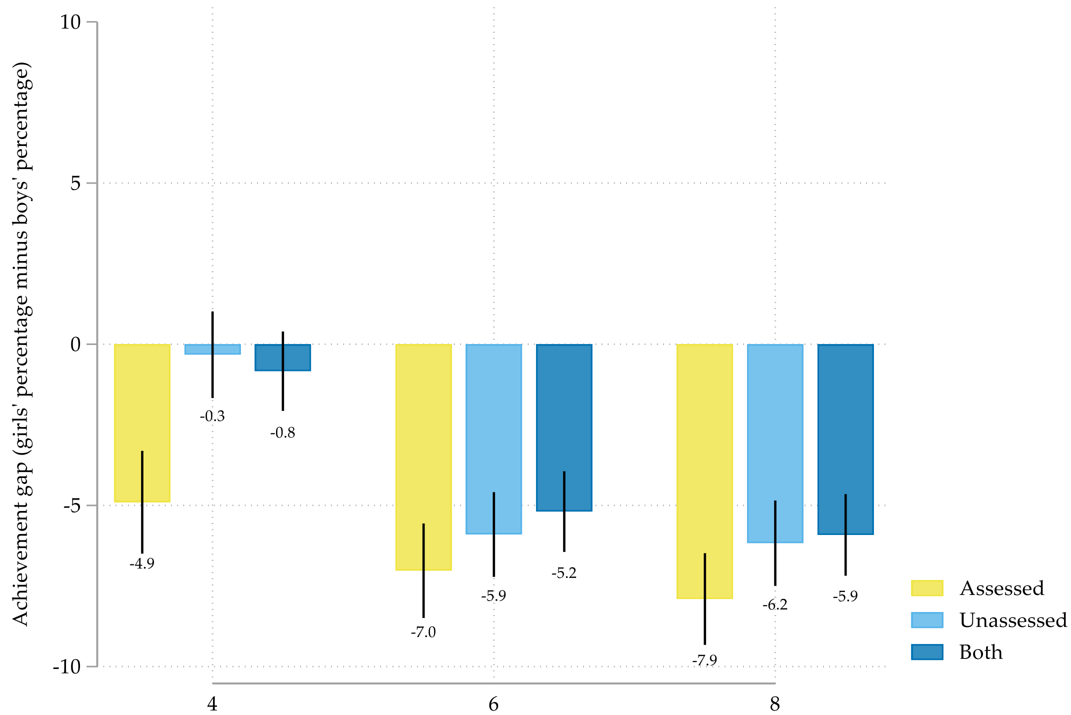
Notes: By students' enrolled grade-level, this figure provides the percentage of students who have mastered the skills on the test, at a fourth-grade level. "Assessed" refers to skills previously assessed by other assessments: number sense and operations. "Unassessed" refers to skills not previously assessed by other assessments: fractions, geometry, and measurement. "Both" refers to both types of skills. Each bar reports on joint mastery of the skills that fall into these three categories. Black vertical bars show 95% confidence intervals.
Source: Authors' elaboration.

Figure 4: Difference in the percentage of students who are proficient in previously assessed and unassessed skills, by sex



Notes: This figure provides information on achievement gaps in the percentage of students who have mastered the skills on the test, at a fourth-grade level (girls' percentage minus boys' percentage). "Assessed" refers to skills previously assessed by other assessments: number sense and operations. "Unassessed" refers to skills not previously assessed by other assessments. "Both" refers to both types of skills: fractions, geometry, and measurement. The three bars to the right report on joint mastery of the skills that fall into these three categories. Black vertical bars show 95% confidence intervals.
Source: Authors' elaboration.

Figure 5: Difference in the percentage of students who are proficient in aggregate proficiency levels, by sex and grade



Notes: By students' enrolled grade-level, this figure provides information on achievement gaps in the percentage of students who have mastered the skills on the test, at a fourth-grade level (girls' percentage minus boys' percentage). "Assessed" refers to skills previously assessed by other assessments: number sense and operations. "Unassessed" refers to skills not previously assessed by other assessments: fractions, geometry, and measurement. "Both" refers to both types of skills. Each bar reports on joint mastery of the skills that fall into these three categories. Black vertical bars show 95% confidence intervals.
Source: Authors' elaboration.

Appendix A: Additional information on the Student Learning Study

The Student Learning Study (SLS) was conceived and executed by Educational Initiatives (EI), a leading assessment firm in India, in 2009. State governments collaborated with the EI study team and provided the necessary permissions to freely conduct the study in the sampled schools and grades. Yet, all study design and field operations fully remained under the oversight of EI. The study was financially supported by Google.org, the charitable arm of Google, but Google.org did not further influence the study design, its data collection, or our analyses.

Appendix Table A.1 provides a comparison of how the SLS differs from other large-scale assessments of mathematics skills. We compare the SLS with the Annual Status of Education Reports (ASER), the National Achievement Survey (NAS), and one example of a randomized controlled trial (RCT) (Muralidharan and Sundararaman, 2011). From Appendix Table A.1, we highlight the SLS' difference from each of the other three assessments.

First, the ASER's sampling strategy differs from the other assessments as it includes students who are not enrolled in schools, and since it tracks students to their homes. However, unfortunately, the ASER does not cover the full range of foundational mathematics skills (omitting spatial skills and geometry, for example). Also, the ASER covers rural locations only. Second, the NAS is the largest study, but there are concerns that government staff (including teachers and principals) may intervene in data collection (compare to Singh, 2020). Moreover, the NAS only started the measurement of foundational skills with its latest round, in 2021, for grades 3 and 5 only, and it does not separately report on students' performance on these foundational skills (vs. at-grade skills). In addition, for grades 3 and 5, publicly available sample test papers do not include questions on spatial skills and geometry. Third, the randomized trial only covered one Indian state (which has since been divided into two states, in 2014). It included below-grade materials in its assessments, but it did not report on results by content domain. Also, it did not separately report on results for those items that capture below-grade (vs. at-grade content) skills.

Table A.1: Comparison with other large-scale assessments from India

	Other assessments			This study
	ASER	NAS	RCT	SLS
Age / Grades (Gr.)	Ages 5-16	Gr. 3, 5, 8, 10	Gr. 1-5	Gr. 4, 6, 8
Comprehensive content domains	No	No	Not reported	Yes
Foundational, below-grade skills	Yes	Limited	Yes	Yes
Coverage	pan-India, rural only	pan-India, urban and rural	1 state, urban and rural	18 states, 1 union territory, urban and rural
Sampling strategy	Selection of districts, sampling of villages, sampling of households, testing all students in the target age range	Selection of districts, sampling of up to 10 schools per district, testing a subsample of up to 30 students per target grade	Selection of 5 districts, sampling of 1 division/district, sampling of 10 mandals/division, sampling of 10 schools/division division, testing all students in target grades	Sampling of districts, sampling of schools, testing all students in target grades
Location	At home	In school	In school	In school
Administration	One-on-one	Group, written	Group, written	Group, written
Invigilation	External	Government, incl. school staff	External	External

Notes: "RCT" refers to Muralidharan and Sundararaman (2011). "Content domains" was rated as "not comprehensive" if tests lack items for entire subdomains of foundational skills, such as spatial skills and geometry, for example. "Limited" indicates that the NAS newly introduced foundational-learning items in its 2021 round, but for grades 3 and 5 only, and without separately reporting on them.

Appendix B: Additional technical details

In this Appendix, we summarize additional technical details of our analytical strategy. We conducted our analyses in five steps.

First, to avoid overfitting and to guarantee that the model development remains independent from the paper's final estimation results, we divided half of our sample into a "training" and the other half into a "holdout" sub-sample. This method is known in the literature as "Random Split Sampling" (see Chen & de la Torre, 2014). We stratified our sub-sampling by state, grade-level, assessment language, and student sex.

Next, using the training data, we identified and screened out items that provided limited information. A student's mastery of a given skill should substantively affect their probability of answering an item correctly (this is known in the literature as item "discrimination"). Students who are proficient on a given skill should also have a higher chance of answering an item correctly, as compared to non-masters (this is known as the "monotonicity assumption"). We removed five items that exhibited low discrimination or violated monotonicity.

In our third step in the process described in section 5, we refined the study's mapping of items to skills (its "Q-matrix"). Our refinement procedure began with a psychometric method proposed by de la Torre and Chiu (2016), which aims to detect mis-specified Q-matrix entries. Then, based on this analysis, we suggested changes to Educational Initiatives' test development team. As a result of this strategy, we modified the item-to-skill mapping for six items.

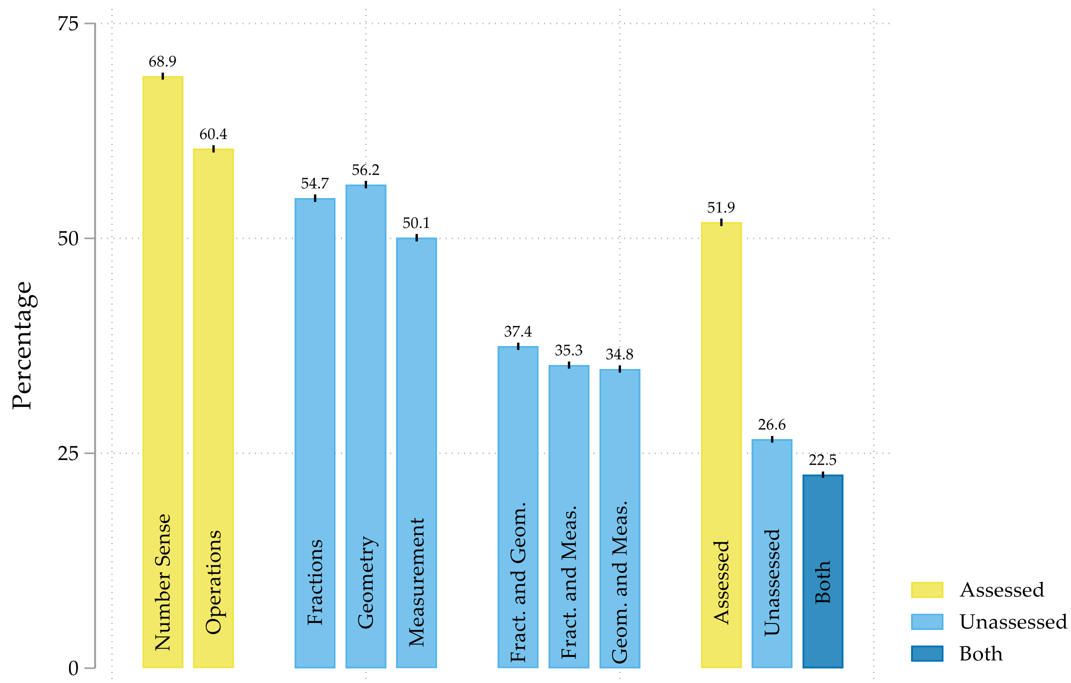
Thereafter, we assessed whether alternative specifications to Equations (1) and (2) could improve our model. We investigated whether the interaction term can be dropped from Equation (1). Following Sorrel et al. (2017) and Ma et al. (2016), our analyses reject dropping the interaction term. We further investigated whether the model could be improved by using a log-linear or logit link, instead of an identity link function. Likelihood ratio tests pointed to the identity link as preferred link function.

Finally, we estimated our model on the holdout sample, fixing all parameters to the training sample's results. Item fit was found to be good when the model is estimated on the holdout sample, as indicated by an average root-mean squared deviation (RMSD) item fit statistic of 0.055. We moreover report on four, common measures of absolute model fit (see Chen et al. 2013). The model fits the holdout data well, given the following fit statistics: a mean of absolute deviations in observed and expected correlations of 0.054, a standardized mean square root of squared residuals of 0.070, a mean of absolute deviations of residual covariances of 0.011, and a mean of absolute values of the centered Q_3 of 0.066.

In terms of reliability, we find that the test's classification consistency is moderate, at the individual level. Following Cui et al. (2012), our calculations suggest an overall consistency of 0.66. This finding is less problematic for the present study as its stated goal is to report on *aggregate* mastery levels. However, we would caution from alternative uses of the same instrument for purposes that require the classification of individual students (e.g., to provide targeted remediation).

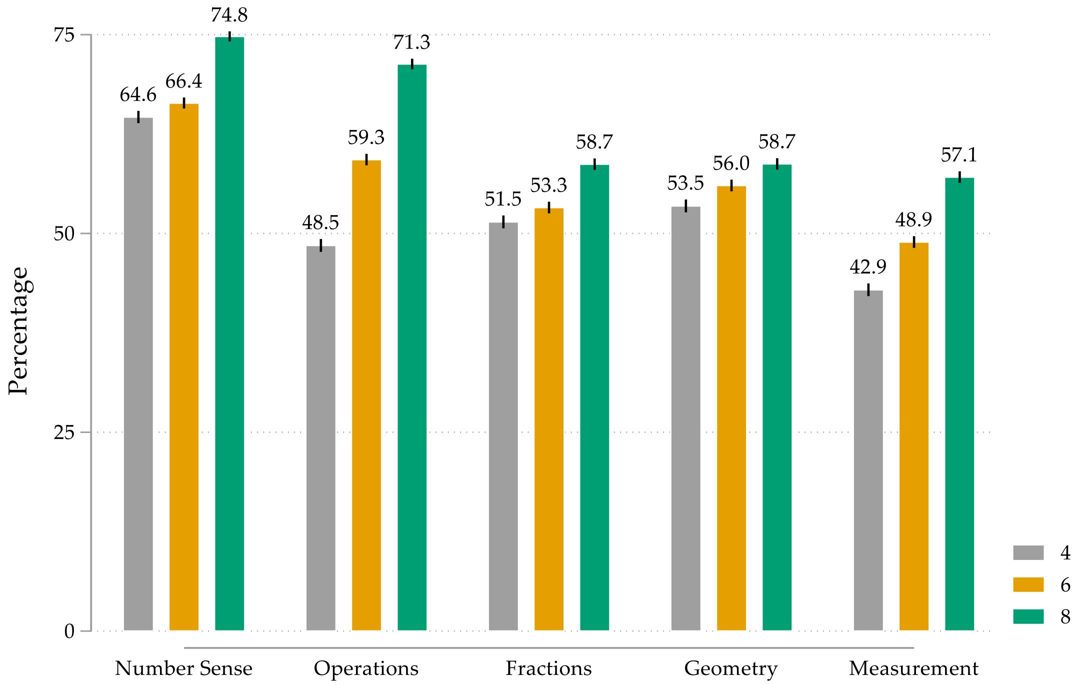
Appendix C: Additional figures and tables

Figure C.1: Percentage of students who are proficient in previously assessed and unassessed skills



Notes: This figure provides the percentage of students who have mastered the skills on the test, at a fourth-grade level. "Assessed" refers to skills previously assessed by other assessments. "Unassessed" refers to skills not previously assessed by other assessments. "Both" refers to both types of skills. The three bars to the right report on joint mastery of the skills that fall into these three categories. "Fract. and Geom." refers to joint mastery of the fractions and geometry skills. "Fract. and Meas." refers to joint mastery of the fractions and measurement skills. "Geom. and Meas." refers to joint mastery of the geometry and measurement skills. Black vertical bars show 95% confidence intervals.
Source: Authors' elaboration.

Figure C.2: Percentage of students who are proficient in previously assessed and unassessed skills, by grade and subskill



Notes: By students' enrolled grade-level, this figure provides the percentage of students who have mastered the skills on the test, at a fourth-grade level. "Assessed" refers to skills previously assessed by other assessments: number sense and operations. "Unassessed" refers to skills not previously assessed by other assessments: fractions, geometry, and measurement. Black vertical bars show 95% confidence intervals.

Table C.1: Number of test questions by grade and content domain

	Fractions and decimals	Geometry	Measure-ment	Number sense	Whole number operations	Anchors	Total
Grade 4							
Up to Grade 4	5	3	5	10	18	14	40
Grade 6							
Up to Grade 4	3	3	1	3	10	14	19
Above Grade 4	4	1	5	5	7	4	22
Grade 8							
Up to Grade 4	1	1	2	3	2	7	9
Above Grade 4	2	3	3	5	6	4	19

Notes: This table displays the number of test questions administered to students enrolled in grades 4, 6, and 8, along with their mapping to content domains. Questions can be mapped to more than one content domain. “Up to”, and “Above Grade 4” refer to the curricular mapping of test questions. “Anchors” refers to test questions also administered to students in at least one of the other two grades. Test questions that are not used in the study are not shown (e.g., grade-8 test questions related to algebra).