

Are You Smarter than a Fourth Grader? New Evidence on the Math Skills of Indian Children*

Andreas de Barros
Harvard University

Alejandro J. Ganimian
New York University

April 02, 2020

Abstract

We leverage a nationally-representative and previously unpublished dataset on the learning outcomes of 101,084 public-school students in grades 4, 6, and 8 across 18 Indian states and one union territory to diagnose their mathematic skills. Importantly, these data allow us to diagnose their achievement not only in commonly assessed skills (number sense and arithmetic) but also on less frequently assessed skills (geometry, fractions and decimals, measurement). We use a novel psychometric approach to estimate the share of students who can meet fourth-grade standards. We find that the arithmetic skills of Indian children may be even lower than previously documented: 52% have mastered frequently assessed skills, but only 27% mastered typically unassessed skills. These children also make less progress than believed: only 22% of fourth graders master geometry, fractions and decimals, and measurement, and the percentage does not change much in grades 6 and 8. Gender gaps in these skills emerge between grades 4 and 6 and persist: while proficiency levels are equal in grade 4 (22%), by grade 6, 25% of girls and 31% of boys show proficiency, and this gap persists in grade 8. Our results indicate that the learning crisis in India may go further than previously diagnosed.

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 the funding provided by the Douglas B. Marshall, Jr. Family Foundation for this project. We especially thank Karthik Muralidharan for his involvement as a collaborator in the early stages of this project and for subsequent discussions. 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 providing excellent research assistance. The usual disclaimers apply. The authors have no conflicting interests to declare.

1. Introduction

India has more school-aged children than any other country in the world. Its school system served 251.3 million students in grades 1 to 12 in 2017 (NIEPA, 2018). In fact, one in five children of primary or secondary school age in the world lives in India (World Bank, 2019). Therefore, to improve children's learning worldwide, it is imperative to address the learning needs of children and youth in India.

Over the past decade, we have learned a great deal about what children and youth in India know and are able to do. Most notably, since 2005, the non-profit Pratham has been conducting surveys of basic skills in reading and arithmetic that are representative of rural areas of the country (Banerji, 2015). The results of this Annual Status of Education Report (ASER) have documented what many analysts regard as a "learning crisis" in the country (Pritchett, 2013). According to the latest figures, even if 97% of children ages 6-14 are enrolled in school, roughly one in two fifth graders cannot read a grade 2 text or solve simple arithmetic operations such as subtractions or divisions (ASER, 2019a).

The ease of administration and focus on basic skills of the ASER tools have played a crucial role in the generation, dissemination, and use of information on student learning in India. They have enabled the ASER Center (the non-profit established in 2008 to conduct the ASER exercise) to mobilize universities, teacher-training institutions, and other community-based organizations to administer the assessments on an annual basis to over half a million children every year (ASER, 2014). They have also served as the primary outcome in impact evaluations of educational interventions in India, helping identify effective programs and policies (e.g., Banerjee, Banerji, Duflo, Glennerster, & Khemani, 2010; Banerjee, Banerji, Duflo, & Walton, 2011b; Banerjee, Cole, Duflo, & Linden, 2007; Duflo, Berry, Mukerji, & Shotland, 2015).¹

The bulk of student assessments in India, however, have focused on whole-number recognition and arithmetic operations, paying little attention to other foundational math skills, such as fractions and decimals, geometry, or measurement.² These other skills matter for children's learning both 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

¹ In fact, it is in great part due to these characteristics that the ASER tools are now used across multiple developing countries in South and East Asia, Sub-Saharan Africa, and Latin America.

² The emphasis on these skills is enshrined in India's Draft Education Policy (see MHRD, 2019).

(National Mathematics Advisory Panel, 2008). It is important that children 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).

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).³

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.⁴ 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.⁵ It cannot be achieved by using other analytical methods based on Item Response Theory (IRT) because there are few items overall for each skill.⁶

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,

³ 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.

⁴ 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.

⁵ 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.

⁶ For example, the fourth-grade assessment includes just five questions mapped to the “fractions and decimals” skill.

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.

Our results provide unique information that can inform current debates on learning levels in India. First, the learning crisis may go farther than previously shown, affecting more skills more severely than number recognition and arithmetic. Second, the low levels of student growth with every additional year of schooling may be even more worrisome than previously reported (see, for example, Muralidharan, Singh, & Ganimian, 2019; Muralidharan & Zieleniak, 2014; Pritchett & Beatty, 2015). Finally, they raise the urgency of addressing inequalities across gender in the country—particularly, given the importance of science, technology, engineering, and math skills in the Indian economy (World Bank, 2018).

The rest of this paper is structured as follows. Section 2 briefly reviews prior research on student learning in India. Section 3 describes the sample. Section 4 presents the dataset that we used for this study. Section 5 discusses our strategy to analyze these data. Section 6 introduces the three main results. Section 7 concludes with a discussion of how our findings could inform policy and practice in India.

2. Prior research

More than a decade of assessments of basic skills have taught us several important facts about children's learning across rural India. First, many children start primary school without being able to recognize letters or numbers. According to the latest ASER report, 43% of children in grade 1 cannot recognize letters and 36% cannot recognize single-digit numbers (ASER, 2019a). Second, most children graduate from primary school without being able to read short texts or perform basic arithmetic operations: 50% of fifth graders cannot read a grade 2 text and 48% 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, 2019a). Third, these two facts have changed little since ASER's inception: in 2005, 42% of first graders could not recognize letters and 58% could not recognize numbers; 60% of fifth graders could not read a text and 27% could not perform operations (ASER, 2006, 2015). Fourth, basic skills vary widely across Indian states. For example, in Kerala and Himachal Pradesh, more than 7 in 10 fifth graders can read a grade 2 text, while only about 2 in 10 of their peers in Jammu and Kashmir and Jharkhand can do so (ASER, 2019a). The reports have also collected impressive contextual information on state school systems.⁷ Recent initiatives to collect similar data through tablets suggests the potential to collect more learning outcomes data more rapidly and more frequently (Gray Matters India, 2018).

These independent assessments have been particularly important in light of the results produced by India's national assessment, which typically offers a much more favorable picture of learning outcomes in the country. The National Council of Education Research and Training (NCERT) regularly administers the National Achievement Survey (NAS) to a sample of more than 2 million students. It assesses students in grades 3, 5, 8 and 10 in math, language, and natural and social sciences on a rotating basis. According to the latest NAS results, 65% of third graders can solve subtractions of three-digit numbers with and without carry-overs and 57% can solve divisions (NCERT, 2014). Importantly, NAS does not focus exclusively on number recognition and arithmetic operations; it also covers domains such as measurement and geometry. However, the fact that the NAS results are consistently more sanguine about the state of learning in the country than both ASER and other domestic studies has raised questions about whether they offer accurate and reliable estimates of children's learning (for an overview of state assessments by 2014, see CIBE, n.d., p. 58).

The proliferation of impact evaluations of education interventions in India has complemented the ASER data by generating a wealth of diagnostic information on the learning levels of children across different states. For example, Banerjee et al. (2007) documented that only 20% of third graders in Vadodara, Gujarat and 34% of those in Mumbai, Maharashtra mastered grade 1 skills in math (e.g., number recognition, counting, and one-digit addition and

⁷ The ASER Center has taken important steps in generating additional information, such as surveying older students (ASER, 2018) and young children (ASER, 2019b).

subtraction).⁸ Similarly, Muralidharan and Sundararaman (2011) were among the first to conduct a representative survey of learning outcomes of primary-school students in an Indian state (Andhra Pradesh), showing that the average student in grades 1-5 could only answer 19% of the multiple-choice questions in their respective test.

In fact, some evaluations have been crucial in providing empirical support for important phenomena previously identified through observations and case studies, such as the discrepancy between curricular expectations and learning outcomes. For example, in their impact evaluation of a computer-assisted learning program in Delhi, Muralidharan et al. (2019) were among the first to show that middle-school students in India perform below grade-level expectations, their achievement increases slower than would be expected, and it varies widely within each grade. Similarly, in their study of working and school children in Kolkata and Delhi, Banerjee et al. (2019) documented how, even when school children were able to competently execute algorithms taught in school to solve arithmetic operations, they were unable to apply those elemental operations to solve real-world problems.

A number of descriptive studies have also helped put learning levels in India in an international perspective. For example, Das and Zajonc (2010) used answers to publicly released items from an international assessment of math and science to show that ninth graders in Orissa and Rajasthan perform below their counterparts in 43 countries that participated in the assessment, and that inequality in learning outcomes in these two Indian states was among the largest in the world. In 2010, Tamil Nadu and Himachal Pradesh participated in an international assessment of math, reading, and science in secondary school, and ranked last among 74 school systems (ACER, 2011).⁹ More recently, Singh (2019) leveraged longitudinal data for four countries to document the differences in the capacity of the school systems in these countries to improve learning outcomes. He found that Indian children already lag behind those in Vietnam—a high-performing developing country—at pre-school age, but that the achievement gap between

⁸ Contemporaneous evaluations in other Indian cities were equally discouraging. For example, Banerjee et al. (2010) found that more than a third (36%) of students ages 7-14 in Jaunpur, Uttar Pradesh could not even read numbers. Similarly, Banerjee, Banerji, Duflo, and Walton (2011a) found that 43% of fifth graders in West Champaran, Bihar and 54% of those in Dehradun and Haridwar, Uttarakhand could not read a grade 2 text. Further, about half of fifth graders in both states made no progress in reading in the following two years.

⁹ It should be noted, however, that problems with the samples precluded those states from obtaining representative information on the achievement of their 15-year-olds (ACER, 2011).

these two countries grows substantially in the first three years of schooling and becomes particularly stark by the end of middle school.

In short, in spite of the impressive progress made in recent decades to document the learning levels and progress of Indian children and youth, several evidence gaps remain. First, nationally representative descriptive studies of learning in math focus on whole-number recognition and arithmetic operations, placing less of an emphasis on other foundational skills, such as fraction and decimals, geometry, and measurement. Second, while some impact evaluations have complemented descriptive studies, their focus on specific cities and states does not allow for a national diagnostic of the skills of children and youth on these less frequently assessed skills. Third, both descriptive and causal studies face important limitations to track the trajectory of children's learning across the school system. Our study seeks to advance existing evidence on all three of these fronts, providing a diagnostic of less frequently assessed skills, doing so on a national scale, and allowing for comparisons across grades.

3. 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.¹⁰ 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).¹¹ 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 grades levels.¹²

¹⁰ 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 students were served by public schools (UNESCO Institute for Statistics 2018).

¹¹ Grade 8 also marks the end of free education, as per India's Right to Education Act (RTE).

¹² The sample in this study represents roughly 72% of the total Indian government school population in these grade-levels.

The sample for the study was representative of the student population in the participating states. EI drew a two-step, stratified cluster sample as follows. First, within each state, it categorized districts by level of development, and it selected two to four districts across those levels (depending on the size of the state), for a total of 48 districts. Then, it 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).¹³ 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 (Educational Initiatives, 2010).¹⁴

4. 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 curriculum and official textbooks in each of the states and union territory in the sample. Then, it pilot-tested the assessments at a small scale and conducted one-on-one interviews with teachers and students to obtain qualitative feedback. Thereafter, it conducted a pilot in 16 schools in three states. The final versions of the assessments were translated into 13 different languages. They were all administered to entire classrooms in a written format in blocks of 120 minutes per subject.¹⁵

A distinguishing feature of this dataset is that EI mapped each question on the tests (also known as “item”, given that they are not always framed as questions) 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). As we discuss in

¹³ For more information on the sampling procedure, see Educational Initiatives (2010).

¹⁴ 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, 2019a; IIPS, 2007).

¹⁵ For more information on test design and administration, as well as copies of all exams and items, see Educational Initiatives (2010).

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 91 unique items: 40 of them were administered in grade 4, 41 in grade 6, and 33 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.

Another important feature of this dataset is that EI also mapped each item to a grade level (based on the curriculum and textbook reviews).¹⁶ 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., 713×24), 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.

5. 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).¹⁷ 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).¹⁸ 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).

¹⁶ We follow Educational Initiatives’ mapping of test questions to grade levels.

¹⁷ We cannot use an Item Response Theory-based approach for our purposes, because of the low number of test questions per domain.

¹⁸ 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.

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. 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 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.¹⁹

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 4.²⁰ 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 $\mathbf{q}_i = (q_{i1}, q_{i2}, q_{i3}, \dots, q_{iK})$ 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

¹⁹ Note that our explanation here is simplified. For additional details, see Rupp, Templin, and Henson (2010).

²⁰ 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.

mastery ($\alpha_{ek} = 1$) or non-mastery ($\alpha_{ek} = 0$) of each math skill $k = 1, \dots, K$ assessed in the test. The vectors \mathbf{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).²¹ In our case, $K = 10$ (five categories of interest and 5 ancillary categories).²²

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

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

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 A.

²¹ 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).

²² 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). 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).

Once we obtain each student’s individual 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 latter. We also compare the percentage of students at this level of mastery across male and female students.

6. 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.²³ Thus,

²³ 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 B.1 in Appendix B.

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.

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 those 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.

7. Conclusion

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, 2019a; 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.²⁴ 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 Indian children and youth. In fact, we believe it may be one of the largest studies addressing this

²⁴ 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.

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.

We conclude with three implications for educational policy. First, the *depth* of the global learning crisis may have been severely underestimated. 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. Our analyses 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 policy makers 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.

References

- ACER. (2011). *PISA 2009+ results: Performance of 15-year-olds in reading, mathematics, and science for 10 additional participants*. Australian Council for Educational Research (ACER). Victoria, Australia.
- 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. (2015). *Trends over time 2006-2014: A supplement to ASER 2014*. ASER Centre. New 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) 2017: Beyond the basics. A survey of youth in rural India*. Retrieved from New Delhi, India:
- ASER. (2019a). *Annual status of education report (ASER) 2018: Provisional*. Retrieved from New Delhi, India:
- ASER. (2019b). *Annual status of education report (ASER) 2019: Early years*. Retrieved from 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. (2011a). *Effective pedagogies and a resistant education system: Experimental evidence on interventions to improve basic skills in rural India*. Unpublished manuscript.
- Banerjee, A. V., Banerji, R., Duflo, E., & Walton, M. (2011b). *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. (2019). *The arithmetic skills of working children are less flexible than previously believed and schooling may not help them improve*. Unpublished manuscript. New York University. New York, NY.
- Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remediating 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.
- Chen, J., & de la Torre, J. (2014). A procedure for diagnostically modeling extant large-scale assessment data: The case of the Programme for International Student Assessment in reading. *Psychology*, 5(18), 1967-1978.

- 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.
- Educational Initiatives. (2010). *Student learning study: Status of student learning across 18 states of India in urban and rural schools*. Ahmedabad, Gujarat: Educational Initiatives (EI).
- Gray Matters India. (2018). *Nudging education: Disrupting the learning ecosystem for masses (First edition)*. ConveGenius Edu Solutions. Noida, Uttar Pradesh.
- 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>
- 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). *Draft national education policy 2019*. 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., & 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.
- NCERT. (2014). *National achievement survey (cycle 3). Class III. Achievement highlights 2014*. Retrieved from New Delhi, India:
- NIEPA. (2018). *U-DISE flash statistics 2016-17*. National Institute of Educational Planning and Administration (NIEPA). New Delhi, India.
- 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.
- 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

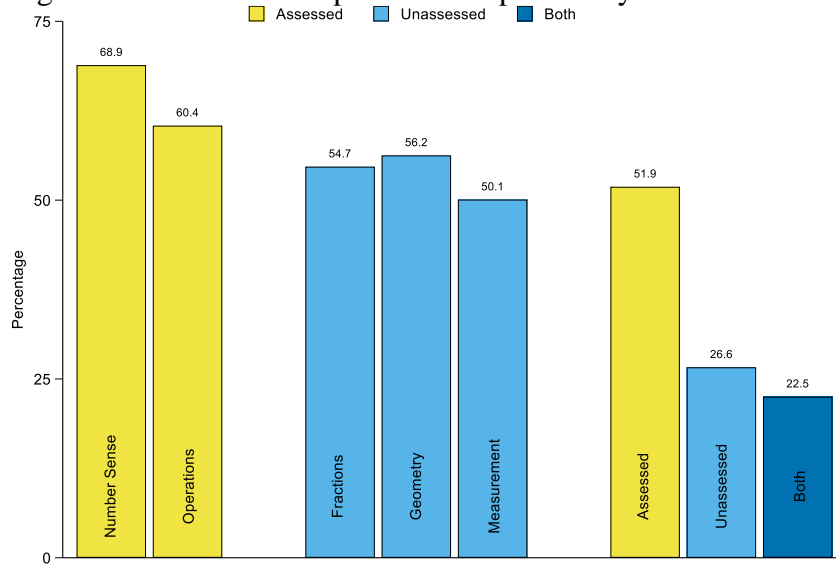
Figure 1: Map of Indian states and union territories participating in the study



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.

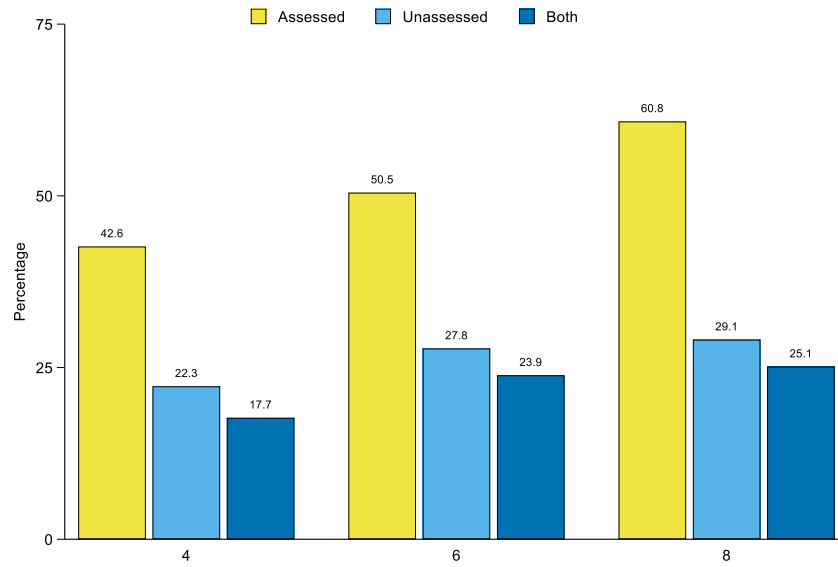
Figure 2: Percentage of students who are proficient in previously assessed and unassessed skills



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.

Source: Authors’ elaboration.

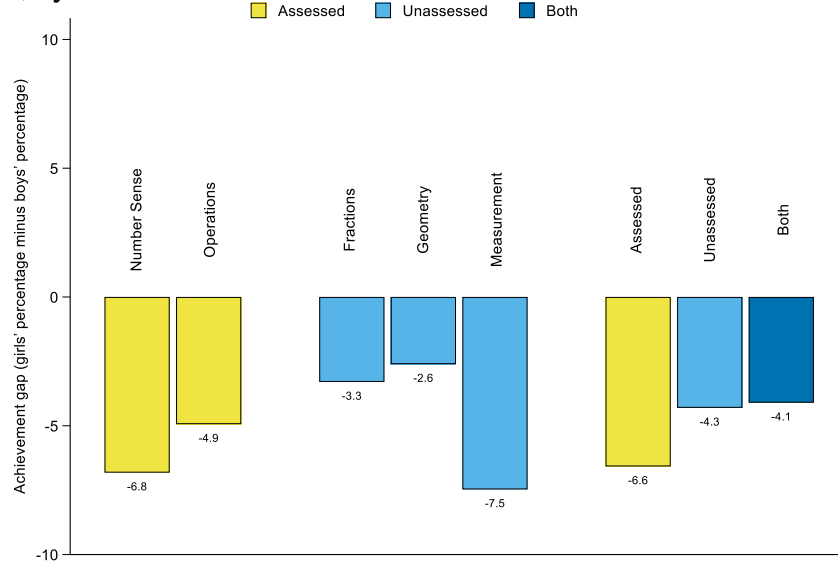
Figure 3: Percentage of students who are proficient in previously assessed and unassessed skills, by grade



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.

Source: Authors' elaboration.

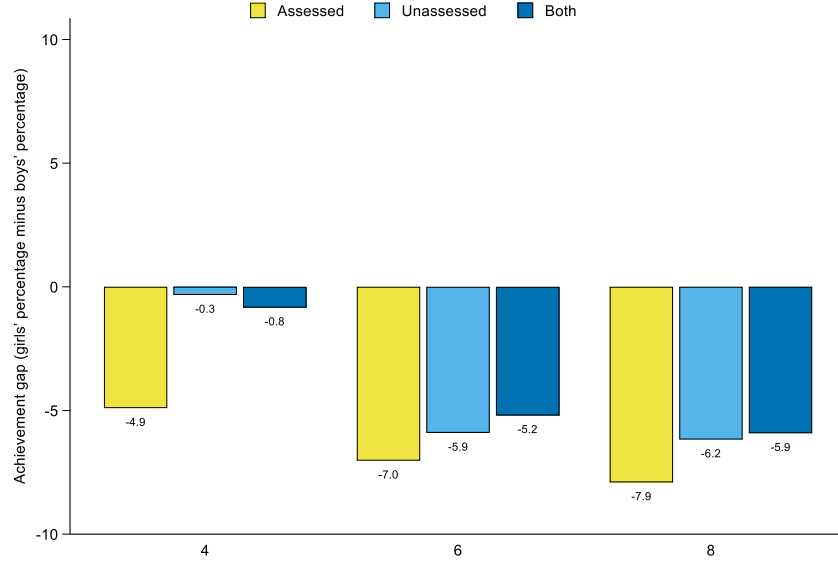
Figure 4: Difference in percentage of students who are proficient in previously assessed and unassessed skills, by sex



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.

Source: Authors' elaboration.

Figure 5: Difference in 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.

Source: Authors’ elaboration.

Appendix A: 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.

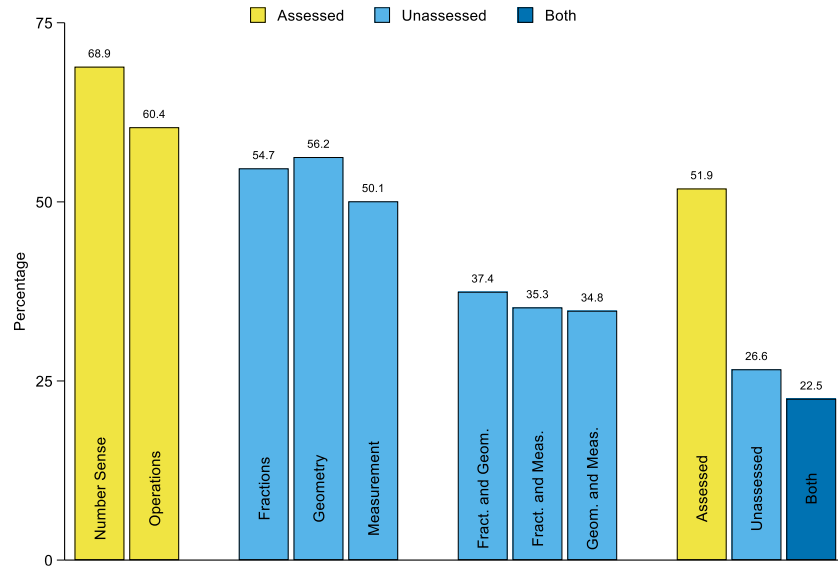
Thereafter, we assessed whether alternative specifications to Equations (1) and (2) could improve our model. We investigated whether the interaction term can be dropped from Equation (1). Following Sorrel et al. (2017) and Ma et al. (2016), our analyses reject dropping the interaction term. We further 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 B: Additional figures

Figure B.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.

Source: Authors’ elaboration.

Authors

Andreas de Barros is a PhD Candidate in Education at Harvard University, at the Graduate School of Education and the Graduate School of Arts and Sciences. His research specializes in program evaluation and evidence-based education policy in less-developed countries.

Alejandro J. Ganimian is an Assistant Professor of Applied Psychology and Economics at New York University's Steinhardt School of Culture, Education, and Human Development. His work focuses on how to reform the incentives, capacity, and supports of school systems in developing countries to improve student learning.