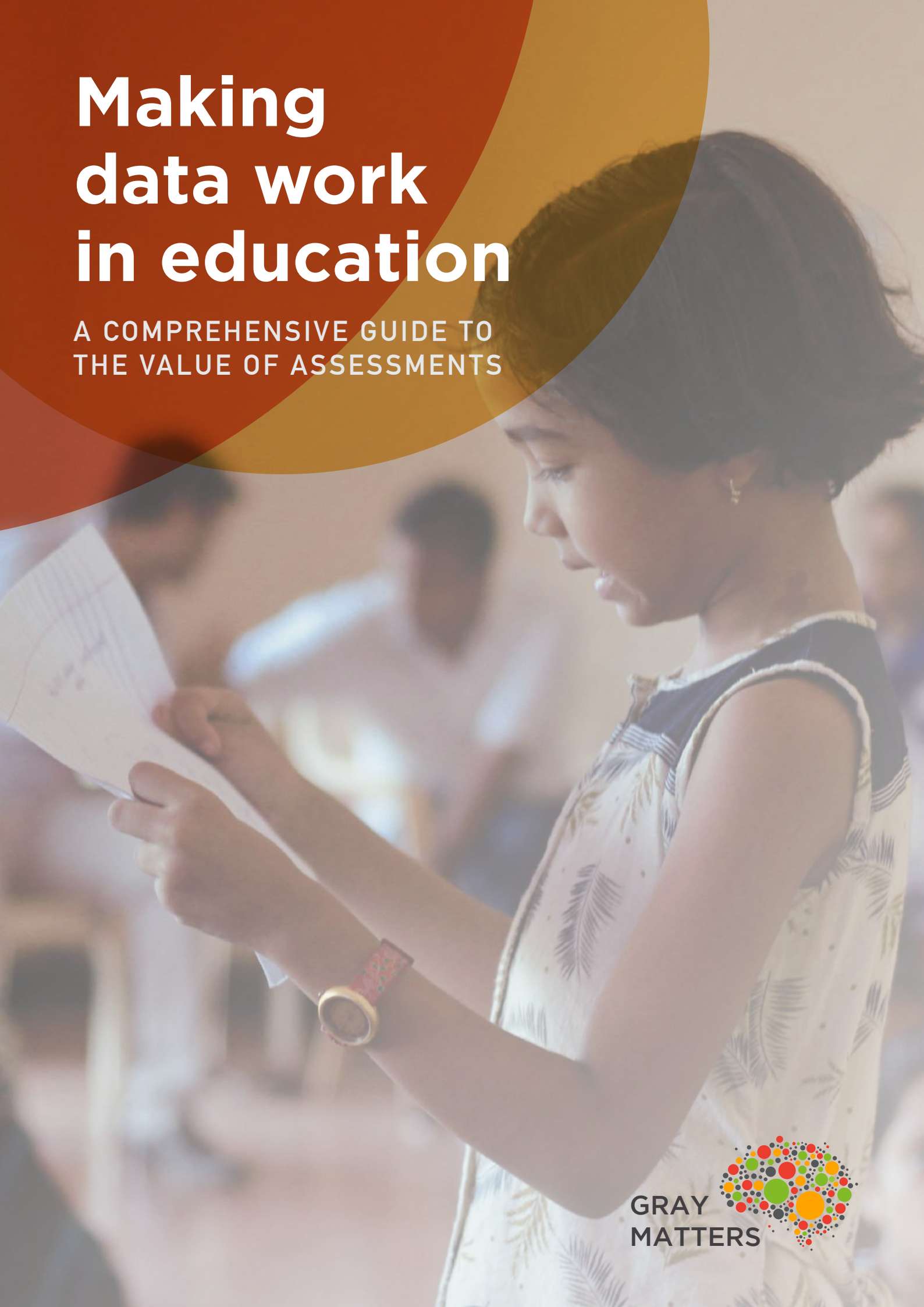


Making data work in education

A COMPREHENSIVE GUIDE TO
THE VALUE OF ASSESSMENTS



No matter what kind of organization you are running, showcasing impact is an indispensable part of the work you do. Now that investors and donors expect quantifiable results, governments, NGOs, businesses, and schools must prove that their programs meaningfully improve student learning outcomes. Educational assessments like GMI's provide value by going beyond what is usually tested in the classroom. Teachers stick to the textbook, but GMI assesses how well students apply skills in practical situations, providing insight into how prepared they are for the future.

What are learning outcomes and why are they important?

Enrolment has gone up in India, but experts have found that school attendance isn't translating into actual learning. Learning outcomes capture how far students have developed skills, like information synthesis, logical reasoning, and critical thinking. This kind of information is critical as a universal metric of impact regardless of what intervention is being implemented, but can be difficult to measure.

So how do we measure student learning outcomes?

The answer to this question is not as simple as it may seem. Let's start with an example.

A teacher gives an exam to her student, Anjali, at the beginning of the school year. Anjali answers 85% of questions correctly. At the end of the year, the teacher gives Anjali another test and she answers 90% of questions correctly.

Have Anjali's skills improved? What if the second test was easier than the first? When the difficulty of the exam itself changes, the new score can't be compared to the old score. We can't know what Anjali has learned over the course of the year.

The percentage of scores give us no information about how many difficult questions were answered correctly versus how many easy questions. And if we can't even compare the two tests Anjali has taken to each other, how are we supposed to compare her scores to students in other schools, districts, and states?

The essential problem is how to compare scores that come from different test instruments.

Tracking Growth with Item Response Theory (IRT)

The most common way of measuring growth in learning outcomes is to compare test scores without knowing the true test difficulty. This method is called “Classical Test Theory” (CTT) and though it is flawed, CTT was the only tool for assessments analysis until statisticians came up with Item Response Theory (IRT) in the 1960s.

IRT starts by doing away with the assumption that every question is equally challenging. Whether Anjali can answer an addition question depends on her skill level and the question difficulty. IRT allows us to determine exactly how difficult the addition question is, which means that we can also determine the skill level Anjali has if she answers correctly.

Additionally, GMI uses questions with gradations of difficulty to precisely determine a student’s learning level, not just above or below one metric of skill measurement. If Anjali can add $2+2$, we know she has basic addition skills. If she can add $2+19$, a problem which requires carry-over into the double digits, we know that she has more advanced addition skills. Carefully designed and piloted questions like these are the foundation blocks of GMI’s statistical analysis.

This approach is the international standard for educational assessments and has been adopted by PISA, ETS, and the NCERT National Achievement Survey. Although IRT is universally endorsed by experts, GMI is one of the only third party education impact evaluators that uses the method in its assessments.

How can GMI’s tools help you?

- GMI’s analysis with IRT solves the problem of comparing results from incomparable test papers, but the model also opens up a range of other benefits that will revolutionize your ability to measure impact.
- Comparing results from different tests with varied difficulty levels allows GMI to track student growth year on year on a single scale. PinAcle, GMI’s signature scale, was built through this model to follow a student’s entire educational journey from Class I to Class X.
- Matching questions with particular abilities gives you more in-depth information, not just on a student’s learning level, but in what particular skills they excel or fall behind in. Questions can be piloted beforehand, evaluated individually, and taken out if they don’t perform well with students, improving the validity and reliability of your data.
- Rigorous and improved test design reduces regional biases between student scores and allows you to reliably assess at any scale, across different regions, boards, and demographic backgrounds.



Interested to know more?
Reach out to Gray Matters.

graymatters.in
contact@graymatters.in
040 49495050