Defining Excellence in Education Through Relevant Benchmarks in India
INTRODUCTION

Continuous and Comprehensive Evaluation (CCE) policy mandated by the ‘The Right of Children to Free and Compulsory Education’ (RTE) Act (2009) underscores the importance of assessments. Though these assessments are being conducted regularly by a large number of schools each year as they adopt CCE policy, educators are rightly concerned that the benefits of such assessments are limited and are not being fully realized. Discussions with educators and NGOs have brought forth multiple issues which limit the potential contribution of assessments in student learning as enumerated below:

• Though assessments happen regularly, each assessment in itself is independent of the others and cannot be linked to any student’s past assessments nor to assessments of his / her peers. Since each assessment is an isolated event, little correlation can be established among the different test results to draw comparisons and meaningful conclusions about the movement in student performances over the years as well as their performance as against their peers.
• Often the objective of the assessment exercise is not clear from the outset or the type of analysis that has to be carried out. There are no rubrics or criteria systematically arrived at to assess students’ performance. Thus, the assessment data that need to be collected are not clearly outlined at the beginning, often resulting in subjective, even superficial assessment, and, data loss or unnecessary data being collected.
• In the absence of any appropriate performance benchmarks or milestones, assessment results often fail to guide future actions/interventions required for students. Given the current state of the students, the assessment results do not provide immediate clarity for setting targets for educators or the intervention programs that need to be enforced in the ensuing 1, 2 or 5 years.

EI, in partnership with Michael & Susan Dell Foundation conducted the research study “Establishing Benchmarks of Student Learning” which tried to address these very issues. This paper discusses ways in which educators can overcome the above challenges by using the results of this study.

ROLE OF PERFORMANCE BENCHMARKS IN EDUCATION

Benchmarking implies comparing a performance to a pre-defined standard in order to evaluate achievement levels. Though some educators object to benchmarking saying that no two students are same and hence should not be compared, the real purpose of benchmarking is not to compare students. Benchmarking should be used as a tool to measure the academic growth of students and design customized curriculums to match each student’s learning needs.

INCREASING EFFICACY OF INTERVENTIONS USING BENCHMARKS

Disconnected assessment in absence of benchmarks

• Decision based on anecdotal evidence and personal experiences
• Effectiveness capped at individual teacher effectiveness
• Subjective and ambiguous learning goals

Benchmarking

• Quantitative understanding of aggregate & subgroup level
• Inferences targeted at student level
• Tracking progress against objective and fair targets

Decision based on evidence of improved interventions
Benchmarking has great relevance in education given that there are very large differences in educational attainment among students not only at a global level but also within India. There are many factors which have contributed to the current state, including but not limited to, diverse backgrounds, funding, varied medium of instruction, accountability of teachers, administration, etc. In such a scenario, it is quite easy to lose sight of the actual objective of education. Benchmarks bring back focus to what is actually important – learning for all students to their full potential. Andreas Schleicher (PISA) summed up the importance of benchmarks in one of his talks where he said, “Local benchmarks are not sufficient. We [tend to] underestimate what is achievable”.

Also, there are many research studies which document the benefits of benchmarks in education. The following are some of the key benefits listed in research literature:

- Fair and objective measure of student achievement (understanding academic achievement in comparison to national standards)
- Effective tracking of student progress and drawing valid comparisons
- Accurate and useful results via a short test which is cost-effective andlogistically easy to conduct
- Research shows that use of standardized tests positively affects student achievement

### STUDY DESIGN AND COVERAGE

Approximately 75,000 students from Government schools (GOV), Low Fee Private schools (APS) and High Fee Private schools (HFP) were covered across India from grades 3 to 7 to capture the entire spectrum of the existing learning levels at the national level.

Government schools were selected from 6 States. These States were selected based on their performance across India in another study conducted by EI in 2009: Student Learning Study. Two States were selected from each of the three buckets:

- states performing above the national average
- states performing at par
- states performing below the national average

Private schools were selected from Tier 1 and Tier 2 cities of India.

Student data were scored using 2PL Item Response Theory model. The Key benefits of using IRT were:

- Estimating item parameters of the questions used in the study which can be used to calculate the learning level of any student in the future and link their score to the benchmark scale;
- Measuring latent ability of students which is a better predictor of learning levels;
- Concurrent calibration (equating) of test scores which allows valid comparison of scores across multiple test papers.

---


3Government schools – Public schools under the purview of state governments, Low Fee Private schools – Unaided private schools charging upto INR 1,000 monthly, High fee private schools – Unaided private schools charging 2,000 or more monthly
**SALIENT FEATURES OF THE STUDY**

**COVERAGE**

<table>
<thead>
<tr>
<th>Class</th>
<th>Government</th>
<th>Low Fee Private</th>
<th>High Fee Private</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>474 Schools</td>
<td>90 Schools</td>
<td>34 Schools</td>
</tr>
</tbody>
</table>

**TOOLS**

70-110 UNIQUE ITEMS per subject

3 TEST FORMS for each subject

**QUESTION TYPES**

- Procedural
- Conceptual
- Application Oriented

**APPLICATIONS OF BENCHMARK SCALE**

As per education literature, a program or policy impact of less than 0.1 SD is typically considered to be a small effect, while more than 0.3 SD is considered large and 0.5 SD very large (Expanding Access and Increasing Student Learning in Post-Primary Education in Developing Countries: A Review of the Evidence, Abhijit Banerjee, Paul Glewwe, Shawn Powers, and Melanie Wasserman).

The scale above is derived by scoring the performance of all the students belonging to different school systems on a common set of test questions. Along with the benchmark scale, item parameters for all the test questions were estimated using IRT. These parameters can be used to link the performance of any student who takes the benchmarking test to the above scale in the future. These scores are referred to as ‘Benchmark scores’ in this paper to distinguish them from the % correct score one gets by using Classical Test Theory.

*Benchmark scales have been developed for classes: 3, 4, 5, 6 and 7, in Language (Hindi, Kannada and Gujarati) and Maths. This is applicable to students from both government and private schools in India. This scale provides meaningful benchmarks and milestones to educators to gauge the current learning levels of students and to design intervention steps customized to each student. The chart below illustrates the benchmark scale (along with a short description of the scale) for class 5, Mathematics.*

**INTRODUCTION TO BENCHMARK SCALE IN THE INDIAN CONTEXT**

Benchmark scales have been developed for classes: 3, 4, 5, 6 and 7, in Language (Hindi, Kannada and Gujarati) and Maths. This is applicable to students from both government and private schools in India. This scale provides meaningful benchmarks and milestones to educators to gauge the current learning levels of students and to design intervention steps customized to each student. The chart below illustrates the benchmark scale (along with a short description of the scale) for class 5, Mathematics.

*Benchmark scores are expressed in terms of effect size, which is a standard metric for measuring the difference in performance between two groups and is expressed in terms of standard deviation. The key benefit of using a benchmark scale is that it allows us to quantify this gap in relative terms.*

The biggest advantage of using benchmark scores is that it allows different test results to be compared across different tests and even across different years or across interventions without compromising on their original objectives. Only a small set of benchmark questions need to be inserted into the test for benchmarking purposes or a small case let can be designed in case no other tests are planned for the students in the academic year. We have illustrated applications in portfolio management.

The scale above is derived by scoring the performance of all the students belonging to different school systems on a common set of test questions. Along with the benchmark scale, item parameters for all the test questions were estimated using IRT. These scores can be used to link the performance of any student who takes the benchmarking test to the above scale in the future. These scores are referred to as ‘Benchmark scores’ in this paper to distinguish them from the % correct score one gets by using Classical Test Theory.

*Mean score is displayed using the ‘Square’ symbol and the distribution has been shown using 1 standard deviation on either side of the mean using the ‘Diamond’ symbol.*

*Please refer to the study “Establishing Benchmarks of Student Learning” for a detailed description of how the scale was generated.*
PERFORMANCE OF DIFFERENT SCHOOL CATEGORIES IN INDIA

Student performance data shows that there is a significant gap between the performance levels of HFP schools with respect to GOV and APS schools. The charts show the performance levels of all the 3 school categories across grades 3 to 7 on the same scale. Learning gaps have been expressed in terms of effect size in order to facilitate comparison across different classes and subjects or even across different studies. Effect size is a standard metric for measuring the difference in performance between two groups and is expressed in terms of standard deviation. The key benefit of using a benchmark scale is that it allows us to quantify this gap objectively and track it across multiple years. To understand the quantum of difference these effect size values represent, they should be viewed/interpreted inline with the 0.3 SD6 effect size which is considered large impact for a successful intervention program in an academic year as per literature.

APPLICATIONS OF BENCHMARK SCALE

The biggest advantage of using benchmark scores is that it allows different test results to be compared across years or across interventions without compromising on their original objectives. Only a small set of benchmarking questions need to be inserted into the test for benchmarking purposes or a small case let can be designed in case no other tests are planned for the students in the academic year. We have illustrated below several benefits which are already being realized by investors and intervention programs using the benchmark scale. In a nutshell, a benchmark scale:

Applications in portfolio management

Benchmarking scale is being successfully used by MSDF to manage their grantees portfolio in the education sector in India. Benchmarking is critical to manage the cycle of - 1) Understanding the current level of the students in an intervention program by benchmarking their performance to national standards, 2) Choosing appropriate intervention strategies which match the current learning levels of the students and, 3) Evaluating the effectiveness of intervention programs by measuring the yearly gains against the desired targets. To be specific, using a benchmark scale to manage portfolio provides the following advantages -

6As per education literature, a program or policy impact of less than 0.1 SD is typically considered to be a small effect, while more than 0.3 SD is considered large and 0.5 SD very large (Expanding Access and Increasing Student Learning in Post-Primary Education in Developing Countries: A Review of the Evidence, Abhijit Banerjee, Paul Glewwe, Shawn Powers, and Melanie Wasserman)
a) Ability to evaluate multiple interventions on the same yardstick

Benchmarking helps to choose appropriate intervention strategies by informing about the current learning levels of students. Simultaneously, it can be used to compare the performance of different intervention programs or strategies. The chart shows that Intervention 1 and Intervention 2 had similar starting point but Intervention 2 managed higher impact on student learning in one academic year. Intervention 3 which had higher learning levels at the baseline, recorded similar learning gains as Intervention 2. The appropriate intervention programs can be selected based on students’ existing needs. Such calculations can also be used to compare the return on investment across different interventions.

b) Tracking performance of state/NGO/Intervention over the years

Benchmarking is useful in tracking the performance on any intervention program across multiple years in order to ascertain the overall direction/effectiveness of the intervention and how much more effort will be required to reach desired targets. The chart above shows the performance of one intervention program, from year 2013 to year 2016, illustrating how benchmarking brings in transparency in evaluating the performance of an intervention program over the years which, in turn, can help portfolio managers/intervention programs in planning optimal resource allocation.

c) Comparison with peers

Benchmarking can also provide comparison with local peers. This comparison can be done using the traditional comparison methods illustrated in the chart below or by using the scaled scores on the benchmark scale. This can be achieved by embedding a few items from the benchmark pool into the State’s assessment paper, thus preventing the need to conduct a separate test for this purpose.

Along with the national standards, a benchmarking scale can also provide comparison with local peers. This comparison can be done using the traditional comparison methods illustrated in the chart below or by using the scaled scores on the benchmark scale. This can be achieved by embedding a few items from the benchmark pool into the State’s assessment paper, thus preventing the need to conduct a separate test for this purpose.

Provides actionable insights to NGOs/Interventions

Beyond the benefits stated above, which apply to both investors and interventions, benchmarks provide granular data at student and question level that can be aggregated in meaningful ways to find actionable insights for interventions. The two charts given below illustrate these points.

Student performance can be broken into buckets and special attention can be accorded to those buckets which consist of higher number of students. Similarly, skill level and question level data can be analyzed to understand the quantum of gap with respect to private schools and can be treated as an interim target to be achieved. These analyses provide quantitative sense of the student learning levels and existing gaps which helps to develop concrete steps to remedy the same.

Strengthening/Making the scale more versatile –

One of the key strengths of a benchmarking scale is that it can be updated periodically and new benchmarks can be added to it as data become available. This will mean that the more the benchmark scale gets used by educators, the more useful it will become in providing deeper and clearer insights.

On-demand standardized test –

The existing item pool and data can be used to provide an on-demand computer based assessment system where any educator can go to the web portal, conduct the test and submit the responses of his students and get immediate results about the performance of his students and their current learning levels vis-à-vis national standards. This can also be scaled to provide deeper and richer analyses on various topics of interest.
Along with the national standards, a benchmarking scale can also provide comparison with local peers. This comparison can be done using the traditional comparison methods illustrated in the chart below or by using the scaled scores on the benchmark scale. This can be achieved by embedding a few items from the benchmark pool into the State’s assessment paper, thus preventing the need to conduct a separate test for this purpose.

**Providing actionable insights to NGOs/Interventions**

Beyond the benefits stated above, which apply to both investors and interventions, benchmarks provide granular data at student and question level that can be aggregated in meaningful ways to find actionable insights for interventions. The two charts given below illustrate these points.

Student performance can be broken into buckets and special attention can be accorded to those buckets which consist of higher number of students. Similarly, skill level and question level data can be analyzed to understand the quantum of gap with respect to private schools and can be treated as an interim target to be achieved. These analyses provide quantitative sense of the student learning levels and existing gaps which helps to develop concrete steps to remedy the same.

**NEXT STEPS**

**Strengthening/Making the scale more versatile** – One of the key strengths of a benchmarking scale is that it can be updated periodically and new benchmarks can be added to it as data become available. This will mean that the more the benchmark scale gets used by educators, the more useful it will become in providing deeper and clearer insights.

**On-demand standardized test** – The existing item pool and data can be used to provide an on-demand computer based assessment system where any educator can go to the web portal, conduct the test and submit the responses of his students and get immediate results about the performance of his students and their current learning levels vis-à-vis national standards. This can also be scaled to provide deeper and richer analyses on various topics of interest.
VOTE OF THANKS

We would like to, specifically, thank the State Governments and all the private schools for giving permissions to freely conduct the study in the schools. We would like to acknowledge the support from the Centre for Civil Society (CCS), ARC, The Teacher Foundation and Dream a Dream for providing data from private schools for sampling of affordable private and high-fee private schools. We also thank the State officials for providing the data from government schools wherever required. We would like to express our gratitude to the Michael & Susan Dell team for their continued support throughout the study.

ABOUT MICHAEL & SUSAN DELL FOUNDATION

Founded in 1999, the central mission of the US-based Michael & Susan Dell Foundation is transforming the lives of children living in urban poverty through better education, family economic stability and health.

The foundation opened its India office in 2006 to foster high-quality education and improve family economic stability among India’s urban poor. Employing a range of financial tools, from traditional philanthropic grants to mission-driven impact investments – the foundation aims to have an immediate and measurable positive impact on the lives of individual children, while at the same time catalyzing systemic change. To those ends, the Michael & Susan Dell Foundation partners with a range of stakeholders, including governments, businesses, NGOs, implementation experts, community organizations and other philanthropies.

Since 2006, the Michael & Susan Dell Foundation has invested more than INR 983 crores in non-profits and social enterprises in India. Learn more at http://www.msdf.org/india

ABOUT EI

Educational Initiatives Pvt. Ltd. (EI), was founded by a group of IIT-IIMA alumni with first-hand experience of setting up and running educational institutions. EI is a social enterprise focused on diagnostic assessments for students and teachers, systemic capacity building of teachers and governments, exploratory research studies that provide inputs on system reforms and recommendations to governments, adaptive and intelligent e-learning systems for personalized learning, and research into how children learn.

RECENT LANDMARK PROJECTS OF EI

Mindspark Centres for Students from Poor Families (2012-2016): EI ran 5 Mindspark Blended Learning Centres in urban slums of Delhi to help children get access to a personalized adaptive learning software to learn Language and Math in Hindi. The program was independently rigorously evaluated by J-PAL’s Dr. Karthik Muralidharan showing encouraging results. The Centres were funded by Central Square Foundation, Tech Mahindra Foundation and Porticus.


Hindi Reading Project (2016-ongoing): This program aims to address the important pedagogic need for an accurate and reliable diagnosis of challenges faced in reading Hindi. The focus of the program is to work with early grade children for identifying reading related challenges. The research is being done to create an electronic diagnostic tool which will be tested and validated with a small group of 370 students in Kota, Jalawar and Bundi districts of Rajasthan.

Capacity Building Workshops for Teachers (2016-ongoing): Capacity building workshops are being carried out for English, Maths and Science teachers of Tamil Nadu. These teachers are being trained on question making, paper setting, sampling and data analysis.

Email – assessment@ei-india.com

October 2016