OUTCOME MEASUREMENT

LANGUAGE AND LEARNING FOUNDATION

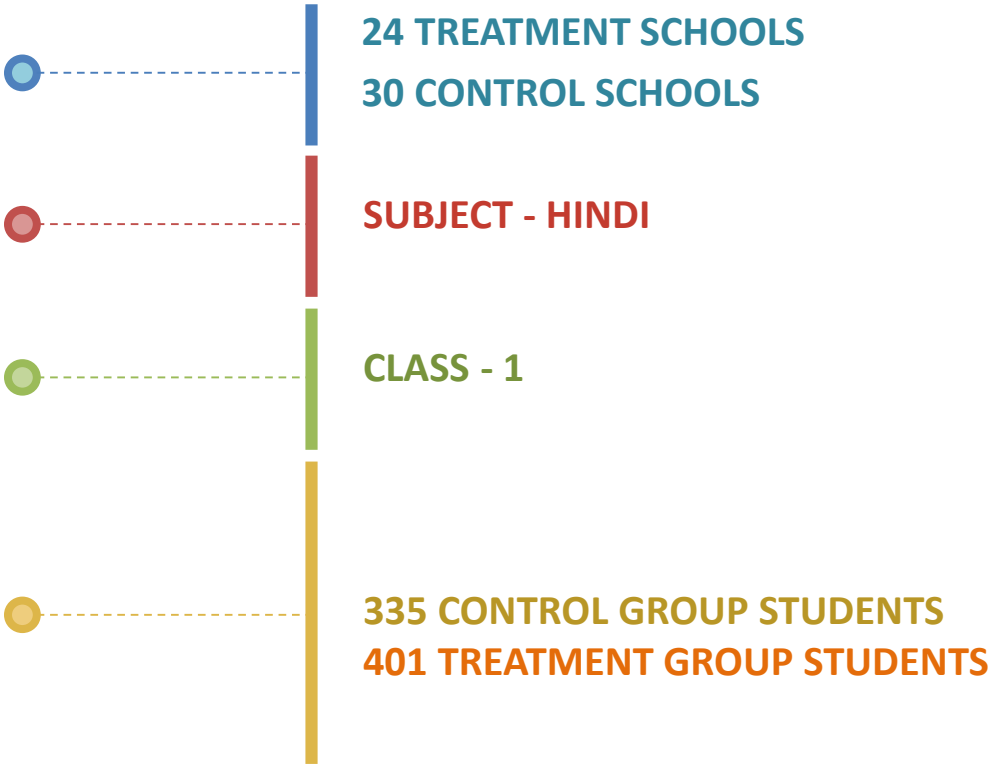
APRIL 2019

GRAY MATTERS INDIA | GRAY MATTERS

INDEX

- ❑ Study Overview
- ❑ GMI Assessment Design Methodology
- ❑ About EARN
- ❑ Student Scale Scores
- ❑ Language Proficiency Levels
- ❑ Student Scale Score Distribution
- ❑ Strand Specific Performance
- ❑ Strand-wise Scale Scores
- ❑ Comparative Performance of Girls vs Boys
- ❑ High & Low Performing Items
- ❑ Appendix

STUDY OVERVIEW



REPORT OVERVIEW

This report presents the findings of the assessments conducted by Gray Matters India (GMI) to measure growth in learning outcome of students. The study assessed students from two groups of schools - schools under the intervention of **Language & Learning Foundation** (Treatment Schools) and schools that are not part of the intervention (Control Schools) to establish the treatment effect.

To achieve this, standardised tests were administered independently in 54 randomly sampled schools. The report represents student scale scores on a vertically integrated scale so that scores across grades are directly comparable along a common continuum. This report describes student scores, growth in scores, student proficiencies and performance by strands.

GMI ASSESSMENT DESIGN METHODOLOGY

1. Defining the Learning Construct

GMI follows Item Response Theory (IRT) and develops assessment instruments which lend themselves to the requirements of IRT modeling. The test design and measurement model are 2-pillars of standardized assessments that are tightly interlinked. GMI has significant know-how and expertise in executing the right assessment design to meet the objectives of the study.

The learning construct defines the learning progression among students in any given subject. It is a detailed blue print of the learning continuum that we expect students to demonstrate. It includes progression in concepts, student abilities and question types. It enables one to divide a measurement scale into performance bands and describe student abilities using standards defined in the construct. The assessment instruments developed by GMI are designed to test the skills and topics illustrated in the schematic, in each subject.

GMI ASSESSMENT DESIGN METHODOLOGY

LANGUAGE	MATH	SCIENCE
Skills Tested	**Skills Tested**	**Skills Tested**

• Letter & word recognition
• Grammar
• Vocabulary
• Reading & Comprehension
  – Information retrieval
  – Interpretation
  – Inference
  – Reflection
– Pronunciation (Spoken)
– Simple sentences (spoken)
– Listening comprehension
	• Numeric problem
	• Paragraph problem
	• Picture problem
	• Multi-step problem
	• Multi-concept problem
		• Conceptual application in experiments
		• Evidence based conclusions
		• Develop hypothesis
		• Interpret scientific data in tabular or picture form

Topics Tested	**Topics Tested**	**Topics Tested**
• Pictures and words • Sentence structure • Narrative passage • Persuasive passage • Procedural passage • Informative passage • Dialogues	• Numbers • Data Interpretation • Geometry • Measurement • Algebra • Statistics	• Biology • Physics • Chemistry • Earth Sciences • Environment • Space Sciences

2. Design Blue Print for Test Instruments

A blueprint is developed for each assessment instrument. The continuum or learning construct is an all-encompassing universal set of skills that is covered in the curriculum. For the purposes of standardized testing within a limited time, core topics and skills that need to be identified, weighted and assigned difficulty levels as per the grade to arrive at a blue print of the test instrument. The blueprint includes total number of questions per test instrument, strands to be covered, topics within each strand and question types. The blueprint also includes specifications for number of equating items required across grades. The test instruments are developed from the item bank according to this blueprint or assessment framework.

3. Development of Test Items

The next step is to develop test items based on the learning construct and the blueprint. Each item has a specific objective on what it is measuring based on the learning construct. The developed items are reviewed by a subject matter expert or panel of experts to ensure appropriateness of the test item, language used, clarity of answer choices and the distractors used in the answer choices. The items are also reviewed to ensure that there is no ambiguity in the way the stimulus is presented or there is no inherent bias in the item based on gender, region, culture etc. GMI follows global best practices in test item development which is listed in the figure below:

GMI ASSESSMENT DESIGN METHODOLOGY

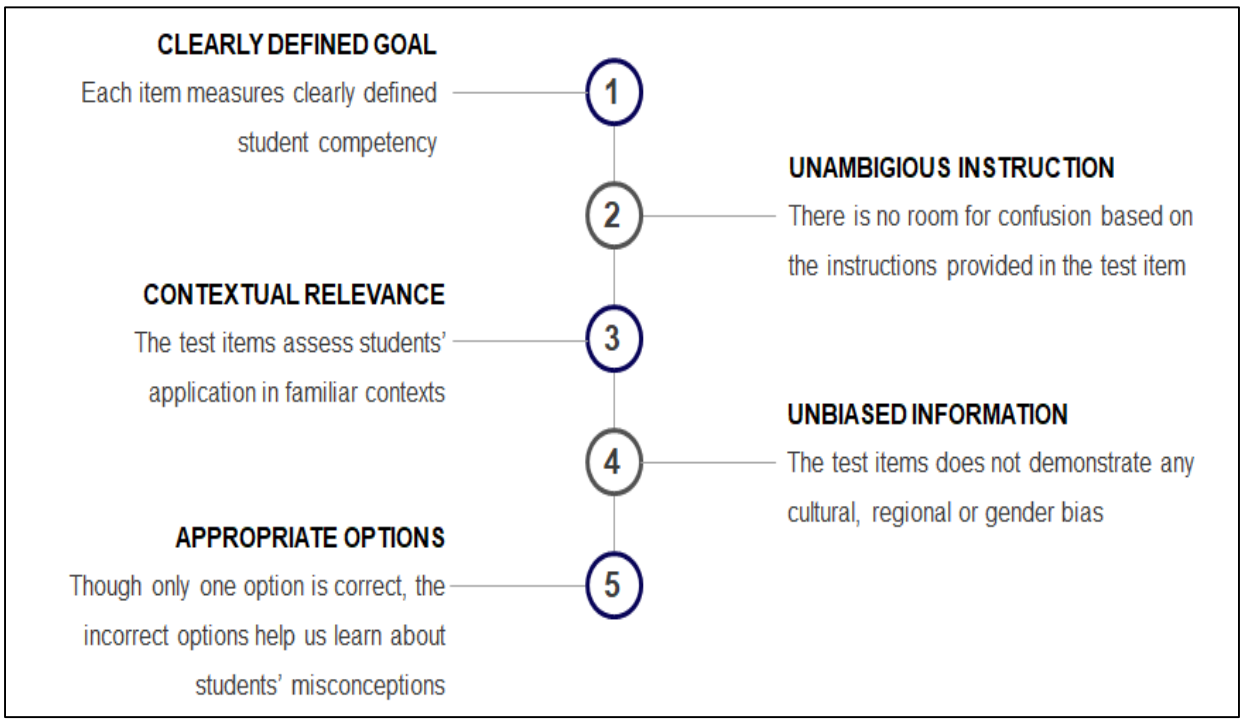


FIGURE:
BEST PRACTICES FOLLOWED BY GMI FOR TEST ITEM DEVELOPMENT

As a function of executing multiple large-scale projects, GMI has a curated item bank consisting of 6000+ items and further items shall be developed to cater to specific requirements that arises during the mapping of the Jharkhand curriculum. The GMI item developers are experienced in developing school-level curriculum-based assessments in Language, Math and Science. Each subject team comprises of resources with expertise and experience across the relevant classes of schooling (classes 1-9).

4. Pilot Test Instruments

The test items are then combined to develop a pilot instrument. The pilot test instruments are used in in respective classes to collect data on how the test instruments and the test items are functioning, while also allowing us to measure the learning level of the students taking this test. In addition to evaluating instrument reliability and validity, the development team conducts on-the-ground observation during the pilot to collect feedback on appropriateness of test items and test instruments to the grade. The test instruments developed for pilots shall have higher number of test items (50% more) allowing GMI to test the suitability of the test items for the respective class and for the test instrument before finalizing it.

GMI ASSESSMENT DESIGN METHODOLOGY

5. Sampling and Test Administration

Probability proportional to size (PPS) sampling is the most frequently used and recommended sampling technique to sample schools. In stage 1, the districts are sampled such that the probability of selection is proportional to the number of students enrolled in the district divided by number of students enrolled in the state. In stage 2, schools are sampled within each selected district, where the probability of selection of each school is proportional to the number of students in the school divided by the number of students enrolled in the district.

GMI is unique amongst assessment companies in India, in having significant on-ground experience in independent test administration. While most agencies take on assessment design, GMI has been heavily involved in conducting assessments of 100-150 schools on a single day and ~700 schools in a span of 2-3 months.

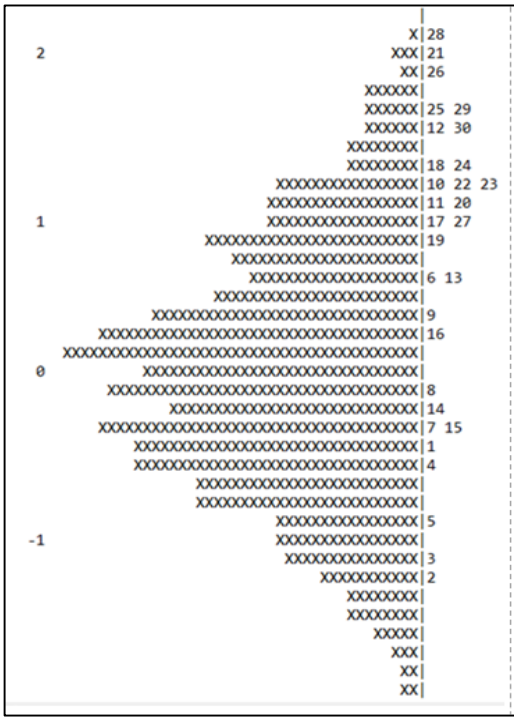
GMI ASSESSMENT DESIGN METHODOLOGY

6. Establish Scale and Benchmarks

Rasch modelling and IRT analysis: GMI uses modern educational measurement methods like Rasch Modeling to analyze and report the results. Student learning outcomes shall be reported using an achievement scale. Students across classes are reported on a single scale and mean comparisons are reported. All comparisons shall be made using mean scale scores and scores in subject specific strands.

This model allows students' ability estimates in an area of learning, to be placed on the same scale as test item difficulty estimates. This means that when test items directly address curriculum learning outcomes, the performance of students can be described in standards statements that relate to student competencies.

GMI ASSESSMENT DESIGN METHODOLOGY



RASCH MODEL & ANALYSIS

Rasch analysis is a psychometric technique that improves the precision with which we construct instruments, monitor instrument quality, and compute respondents' performances

01 TEST ITEM FIT

It is a correlation between the student performance in the item and in the test. It tests the item construct and the concept underlying the test item.

02 TEST ITEM VALIDITY

Differential item functioning is a potential source of item bias.
Eg: Gender bias, Rural/Urban bias etc

03 TEST RELIABILITY

It is a measure of how reliable is the measurement of a person's ability.

04 DIFFICULTY ESTIMATES

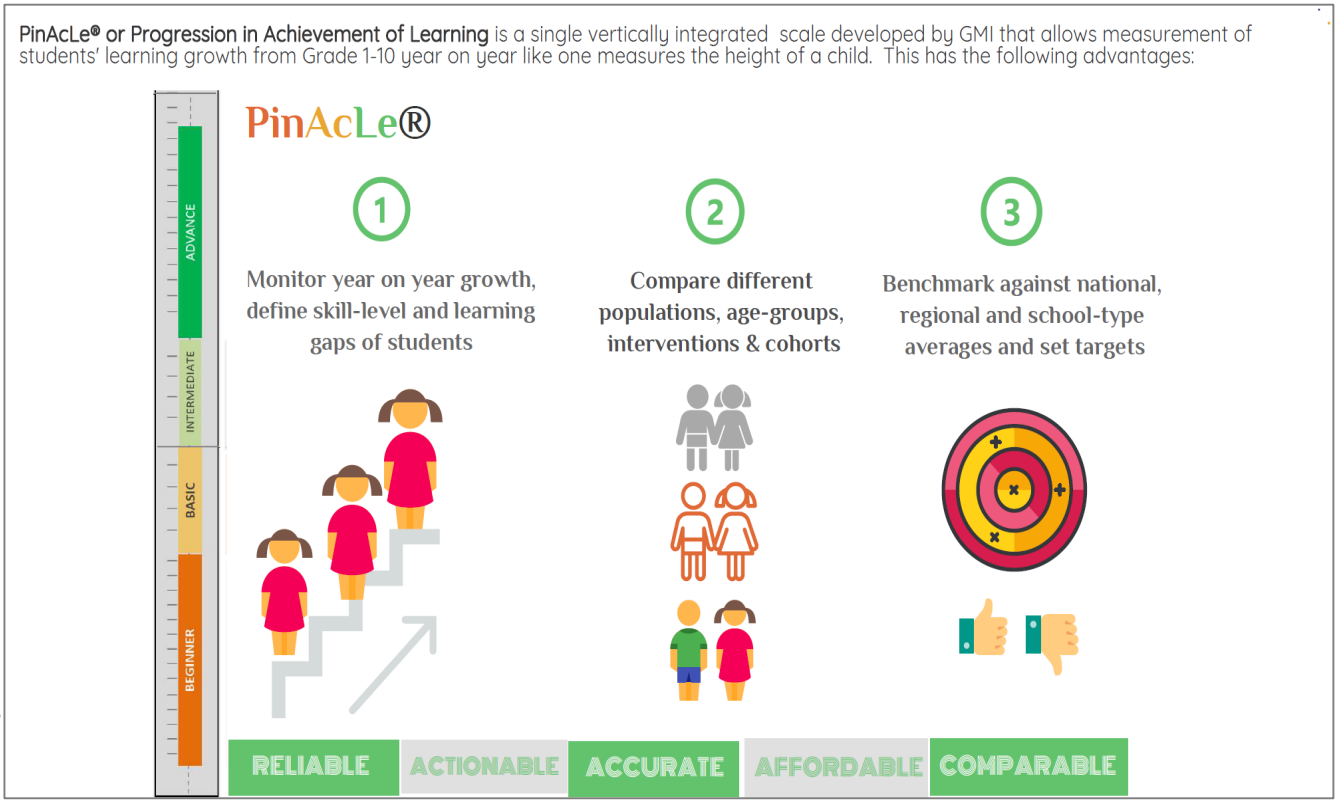
Item difficulty is estimated in the Rasch model by analyzing student responses from the sample.

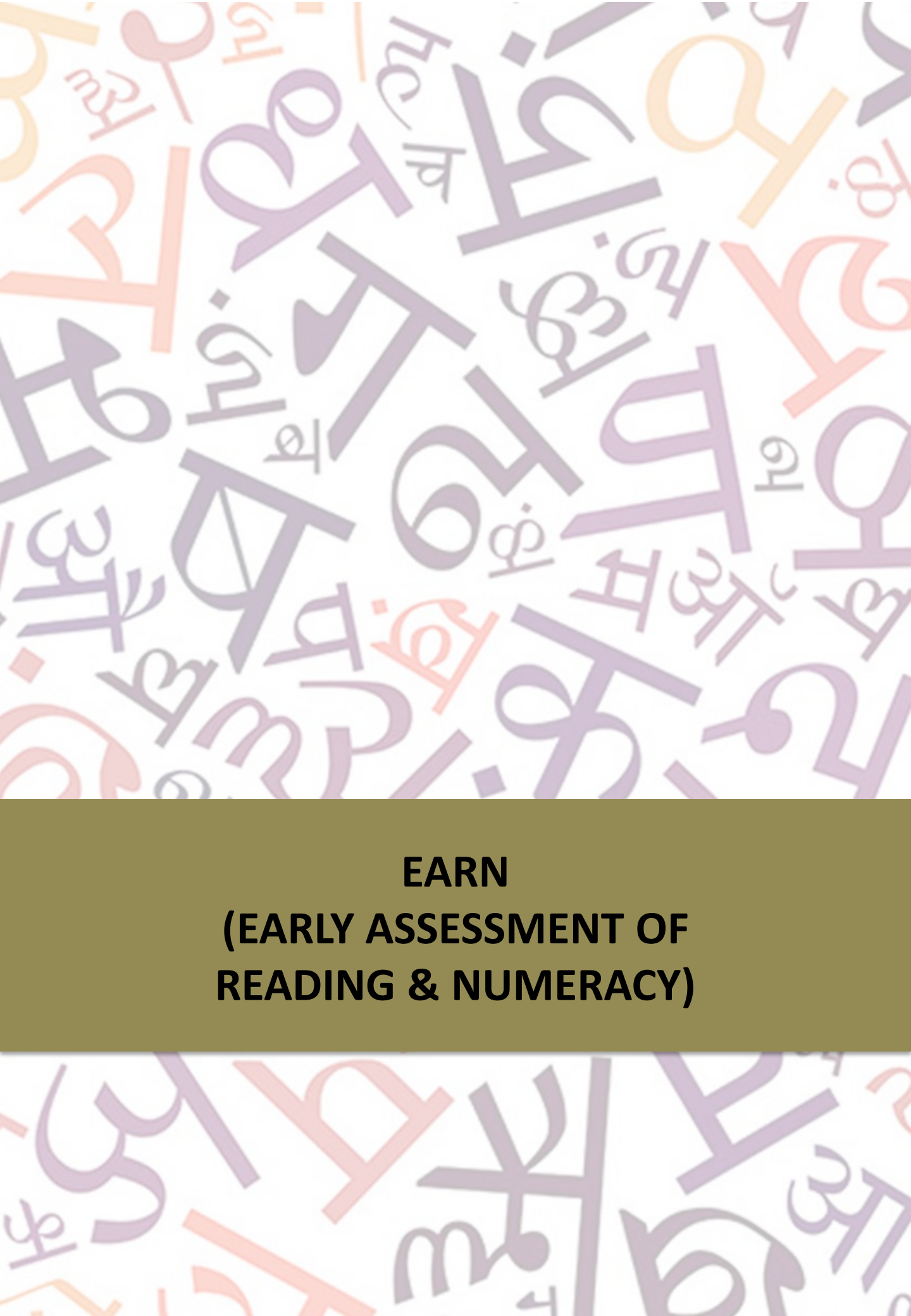
This model allows students' ability estimates in an area of learning, to be placed on the same scale as test item difficulty estimates. This means that when test items directly address curriculum learning outcomes, the performance of students can be described in standards statements that relate to student competencies

Panel of subject matter experts to determine grade-level cut-offs: To establish grade-level cut-offs or to set "standards" a combinatorial approach of quantitative methods above is used along with qualitative methods described below. GMI proposes the use of a combination approaches of Angoff and Bookmark methods to arrive at the qualitative cut-off for establishing grade-level.

7. Final Scale, Benchmarks and Reporting Metrics

GMI has developed a proprietary scale, PinAcLe® (Progression in Achievement of Learning) that reports scores of all classes on a single vertically integrated scale. This allows one to measure growth in learning outcome between classes 1-10. The scale can be divided into performance Levels. Each Level shall have standard statements to describe student abilities at that level.





EARN
(EARLY ASSESSMENT OF READING & NUMERACY)

ABOUT EARN

1. **Gray Matters Early Childhood Assessment (EARN)**
    - A diagnostic assessment program that focuses on aspects of early childhood skills essential for school success
    - Tracks child's progress in meeting key developmental milestones in the areas of Language, READING & Mathematical Skills
    - Based on Early Grade Reading & Mathematics Assessment Framework, contextualised to NCERT guidelines
2. **How EARN helps Schools & Teachers**
    I. EARN provides critical information regarding each student, including:
        - Does the student have prerequisite skills to master the material to be taught?
        - What is the student's current level of knowledge and skills?
        - Could the student be gifted/talented or have learning difficulties that may require further testing?
        - Has the child acquired sufficient skills during the academic year? What is the growth achieved?

II. EARN provides a baseline measure against which student progress can be measured throughout the year. This can help teachers in adjusting instruction and creating individualised learning plans for students. In the endline, the growth in learning of a child is tracked

3. Features of EARN

I. Diagnostic test

- EARN is a diagnostic test that helps educators understand what students know, what they are yet to learn and what they have learnt by the end of academic year – The test will therefore progress from below grade-level skills to above grade-level skills.
- EARN is a not a test of achievement or competitiveness - it tests a continuum of skills against which children are mapped.
- The test is designed as a continuum from easy to difficult questions and testing maybe terminated (as per pre-set instructions) if child is unable to answer a series of questions

ABOUT EARN

II. Standardized test:

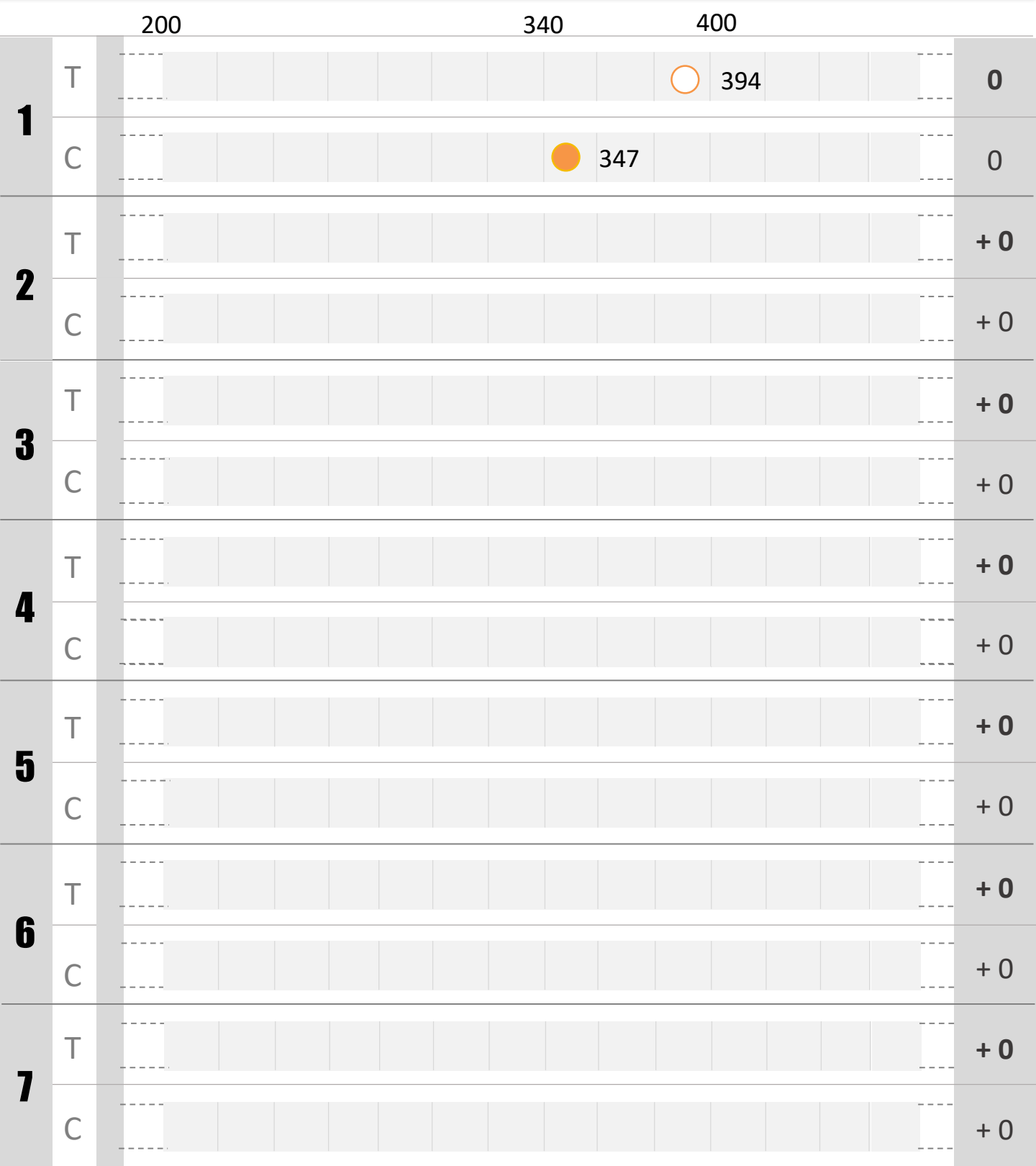
- Standardization in test administration and data capture to ensure reliability of assessment program. This involves a standard amount of time allocated for each activity and data capture on digital devices.
- Overall test is designed for efficient, consistent and objective grading of responses to provide , accurate and easy to interpret reports that helps teachers make informed instructional decisions
- EARN is different from school test in that it is meant to be an objective observation by a 3rd party – while assessors will be encouraging, they will not prompt/help a child outside the pre-set process.

4. **Sub Skills tested for this study:**
    - Picture Recognition
    - Picture Matching
    - Sentence Repeating
    - Listening Comprehension
    - Picture Description
    - Sequence Identification
    - Letter Reading
    - Word Reading
    - Reading Comprehension
    - Dictation
    - Picture Description

RESULTS OF HINDI LANGUAGE COMPETENCY ASSESSMENT

STUDENTS SCALE SCORES (TREATMENT vs CONTROL)

HINDI



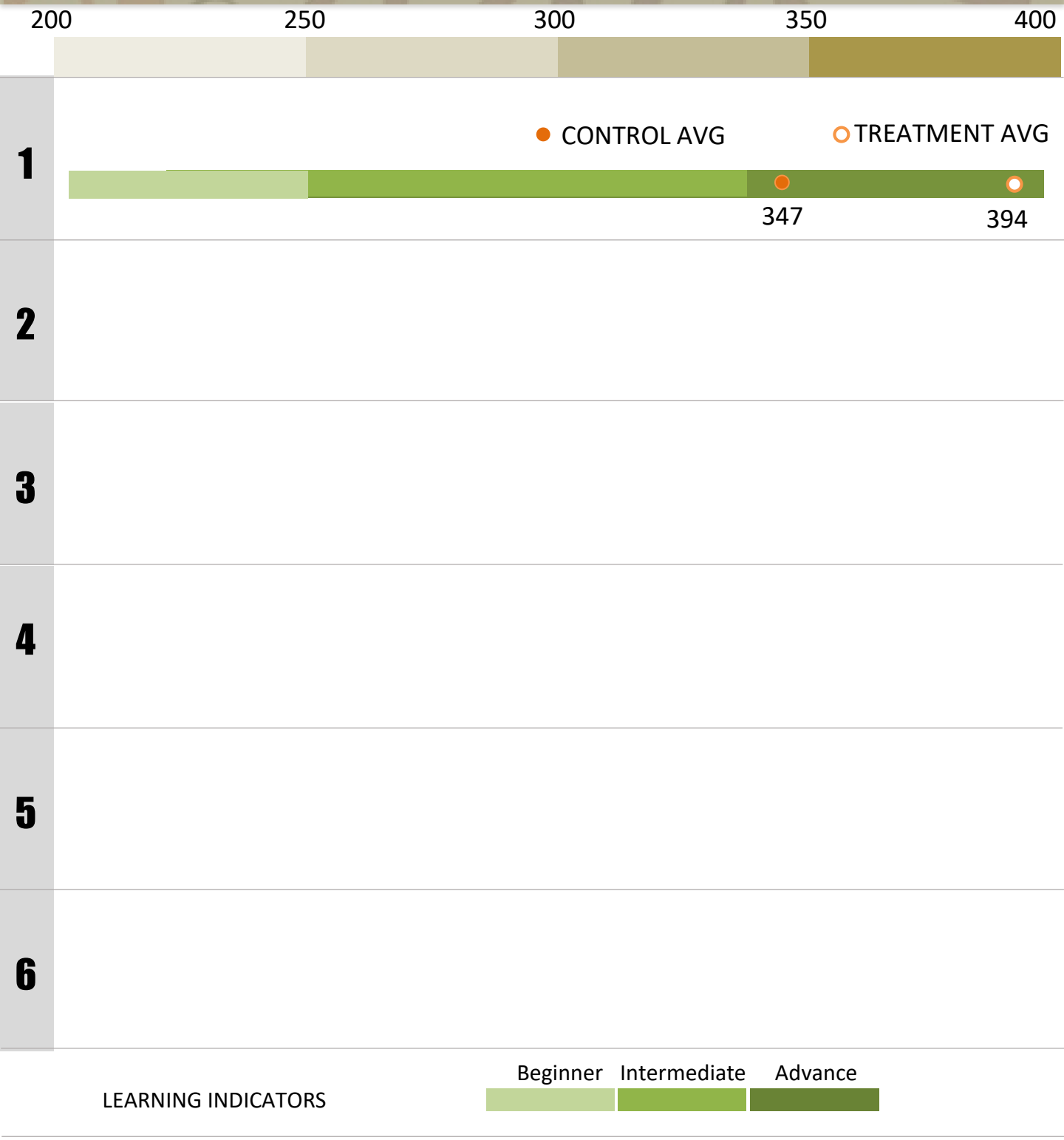
T: Treatment C: Control

LANGUAGE PROFICIENCY LEVELS

GRADE 1

ADVANCE (>= 330)	**Students decode alphabets and sight words by associating them with appropriate sounds. They interpret and link aspects of their daily life when speaking.** Students identify alphabets (vowels and consonants) and recognise some 2-letter sight words. They read simple subject-verb-object sentences and match them with the appropriate picture. When presented with pictures they speak by correlating implicit sequence of events and linking it with their daily life.
INTERMEDIATE (270-329)	**Students develop awareness of unfamiliar things and identify them. They start decoding alphabets by associating them with appropriate sounds. They listen with comprehension and can speak in complete structured sentences.** Students identify pictures of wild animals, birds, furniture, places, etc. by their names. They start identifying some high-frequency alphabets (mostly vowels). They answer direct questions after listening to 1-2 sentences that are read out to them. They describe a series of pictures presented to them in complete sentences.
BEGINNER (<270)	**Students have awareness of everyday things and identify them by names. They listen with the intent to repeat and can speak in partial sentences to express themselves.** Students identify pictures of fruits, vegetables, domestic animals, etc. by their names. They can listen to short sentences or phrases and repeat them verbatim. They describe events in a picture using partial sentence or phrases.

STUDENTS SCALE SCORE DISTRIBUTION ACROSS PROFICIENCY LEVELS

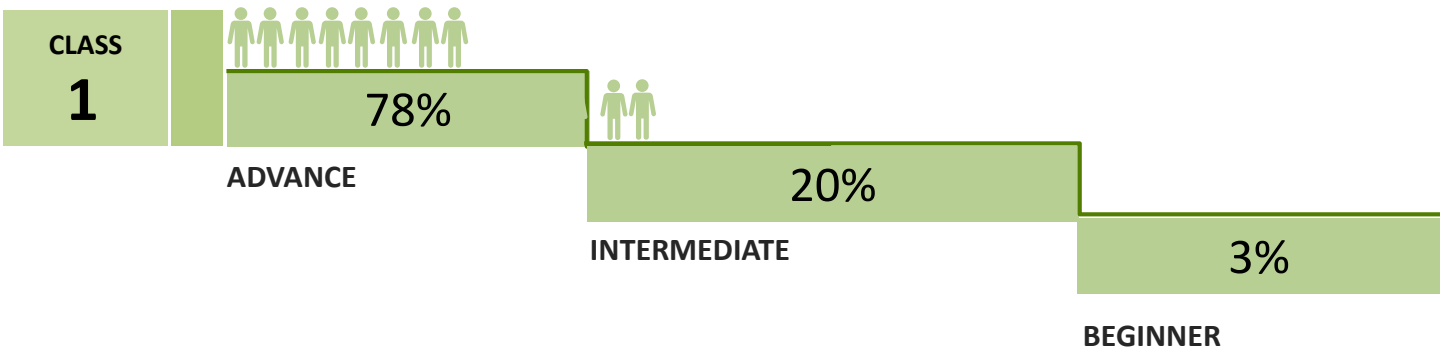


The slide shows the Average Scaled Scores for Control and Treatment group and their respective position on proficiency band. Treatment group is preforming better compared to control group

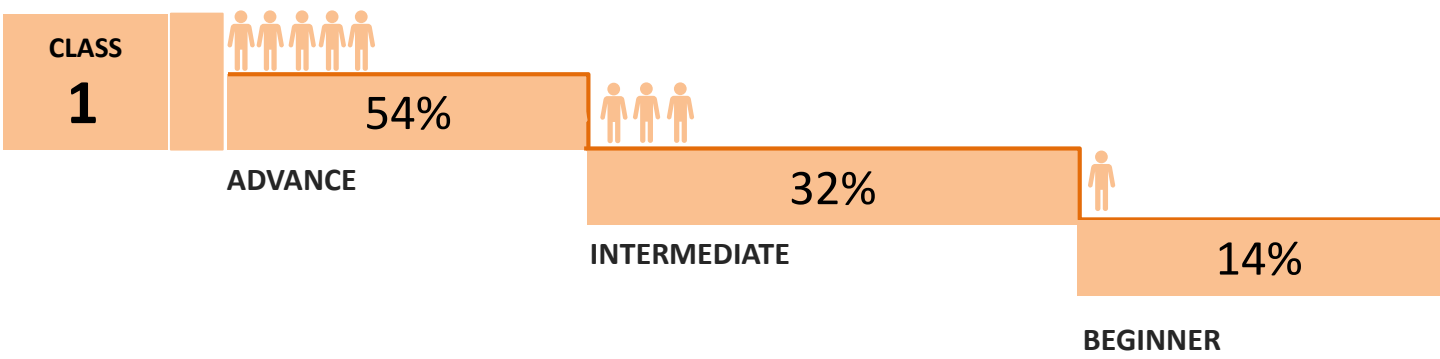
STUDENTS PROFICIENCY LEVELS
DISTRIBUTION BY PROFICIENCY LEVELS

HINDI

TREATMENT GROUP



CONTROL GROUP



The slide shows the percentage of students at different proficiency levels across control and treatment groups. Majority of the treatment population is advanced level.

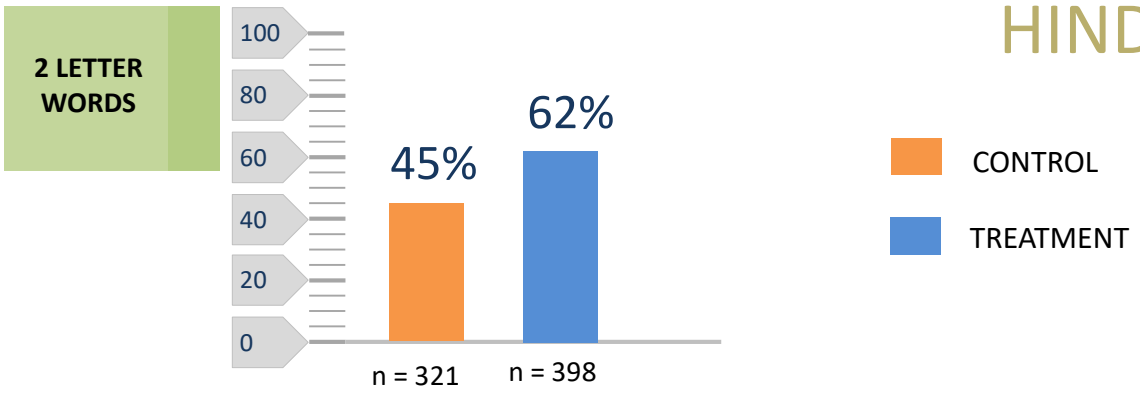
CLASS 1 – HINDI: STRAND SPECIFIC PERFORMANCE

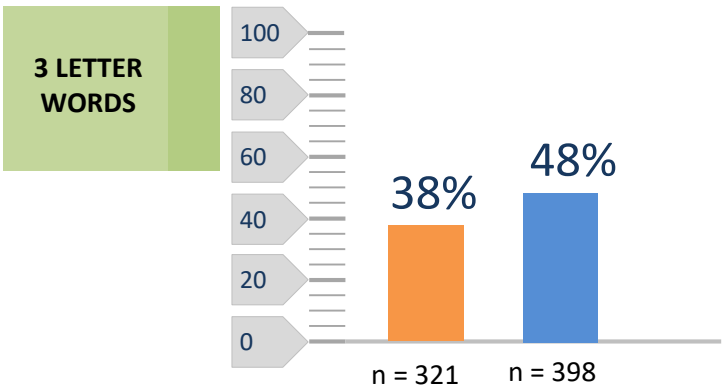
Strand	CONTROL	TREATMENT
Picture Recognition	92%	93%
Picture Matching	65%	81%
Sentence Repeating	93%	99%
Listening Comprehension	78%	85%
Picture Description	83%	88%
Sequence Identification	27%	26%
Letter Reading	73%	90%
Word Reading	44%	70%
Reading Comprehension	30%	48%
Dictation	36%	51%
Picture Description	52%	58%

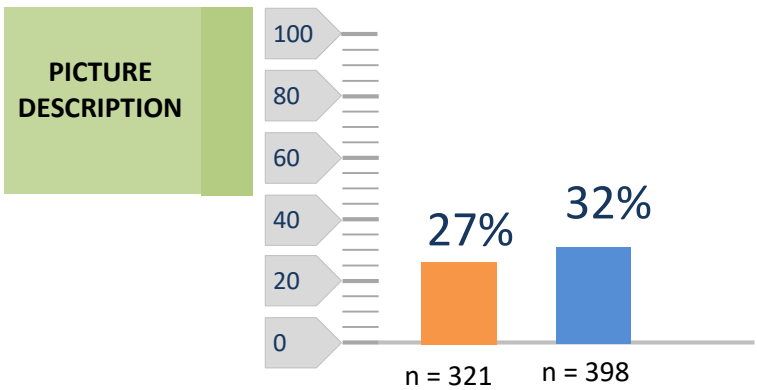
The slide shows the percentage of students answering correctly across Skills assessed for Treatment and Control group.
Treatment group students have performed better in all the strands except for sequence identification.

COMPARATIVE PERFORMANCE OF WRITING SKILLS

HINDI







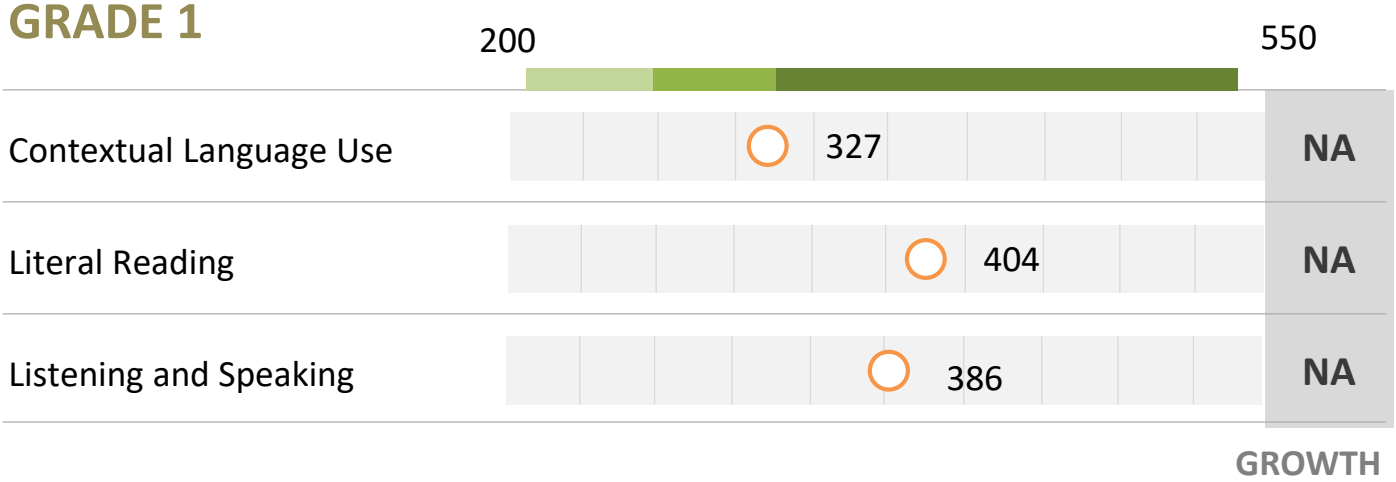
Children can write 2/3 letters words where as descriptive writing needs attention.

STUDENTS SCALE SCORES STRAND WISE*

HINDI

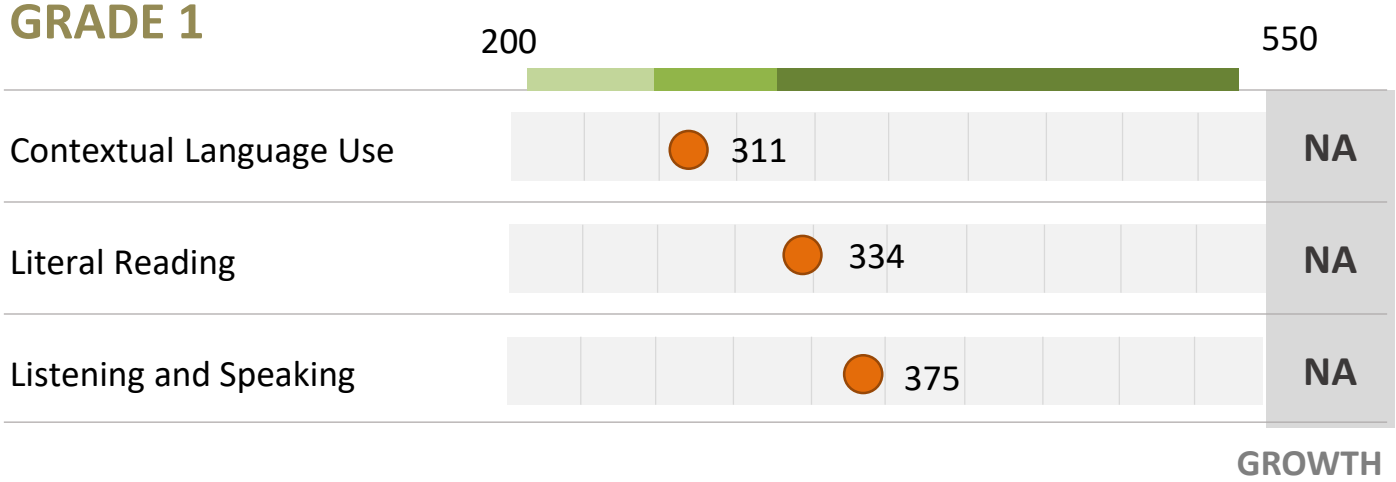
TREATMENT GROUP

GRADE 1



CONTROL GROUP

GRADE 1

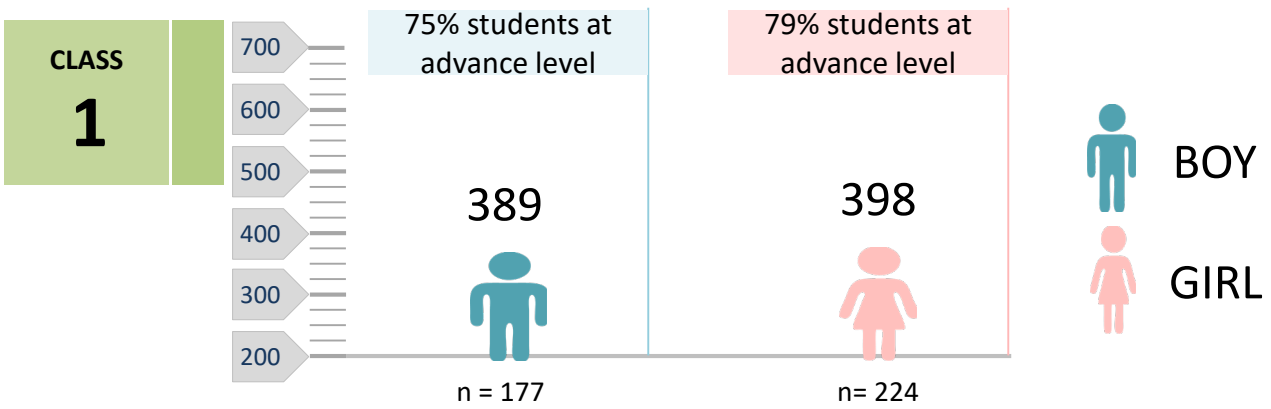


** Strands defined corresponding to various EARN Assessments conducted across the country.*

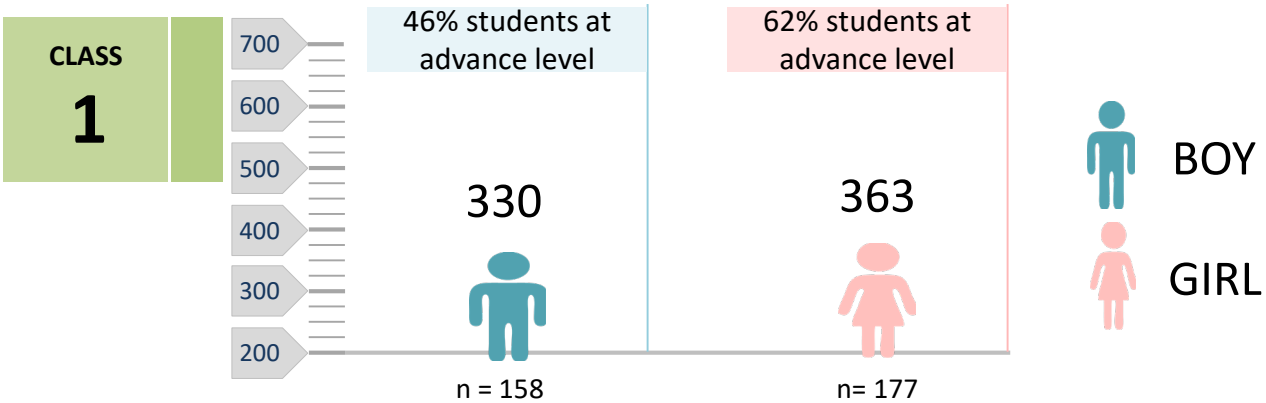
COMPARATIVE PERFORMANCE OF BOYS vs GIRLS

HINDI

TREATMENT GROUP



CONTROL GROUP



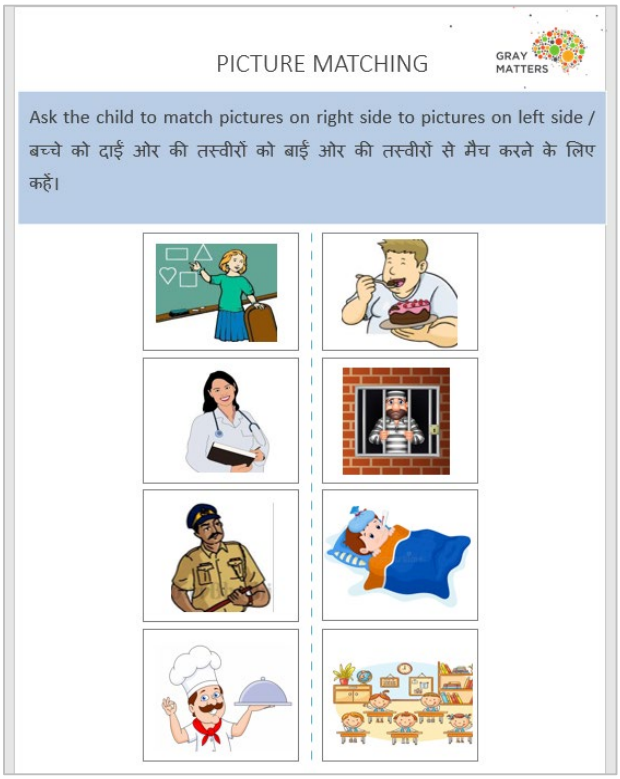
HIGH & LOW PERFORMING ITEMS
ITEM SPECIFIC PERFORMANCE

HIGH & LOW PERFORMING ITEMS
1. CONTEXTUAL LANGUAGE USE

Identify an object by its function

Percentage Correct
Treatment- 99%
Control- 97%

Match profession to place of work - Doctor : Patient

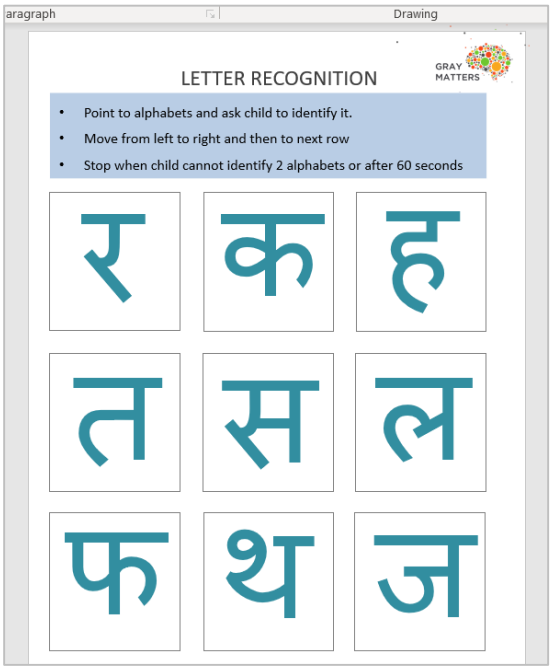


Percentage Correct
Treatment- 78%
Control- 69%

Note: Items are confidential – Do not share outside this discussion group.

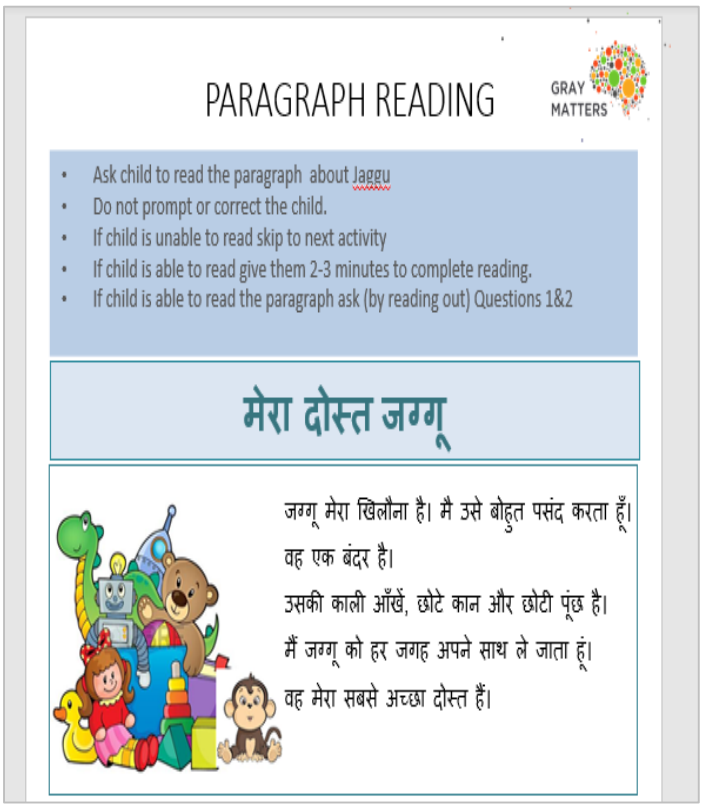
HIGH & LOW PERFORMING ITEMS
2. LITERAL READING

Identify the alphabet



Percentage Correct
Treatment- 98%
Control- 80%

Read 1st paragraph (4 lines) from the short narrative text

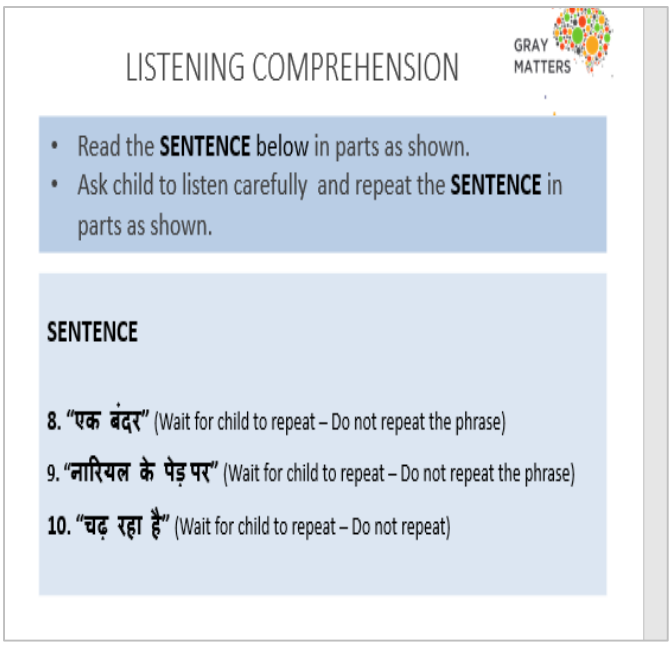


Percentage Correct
Treatment- 52%
Control- 41%

Note: Items are confidential – Do not share outside this discussion group.

HIGH & LOW PERFORMING ITEMS
3. LISTENING & SPEAKING

Listen and repeat a phrase from a sentence



Percentage Correct
Treatment- 99%
Control- 94%

Describe the series of pictures in order

Percentage Correct
Treatment- 15%
Control- 21%

Note: Items are confidential – Do not share outside this discussion group.

APPENDIX

CLASS 1 – HINDI: ITEM SPECIFIC PERFORMANCE (1/3)

Item descriptor	Strand	CONTROL	TREAT-MENT
Identify an object by its colour	Picture Recognition	**83%**	**83%**
Identify an object by its function	Picture Recognition	**97%**	**99%**
Identify an object based on category - Fruit	Picture Recognition	**95%**	**98%**
Match pictures of profession to place of work - Teacher : Classroom	Picture Matching	**69%**	**81%**
Match profession to place of work - Doctor : Patient	Picture Matching	**69%**	**78%**
Match profession to place of work - Police : Prisoner	Picture Matching	**75%**	**88%**
Match profession to place of work - Cook : Person eating	Picture Matching	**67%**	**81%**
Listen and repeat a phrase from a sentence - Subject	Sentence Repeating	**94%**	**99%**
Listen and repeat a phrase from a sentence - Object	Sentence Repeating	**93%**	**99%**
Listen and repeat a phrase from a sentence - Verb	Sentence Repeating	**94%**	**98%**
Listen to a short story (2 lines) and answer a simple 'who' question	Listening Comprehension	**77%**	**86%**
Listen to a short story (2 lines) and answer a simple 'what' question	Listening Comprehension	**80%**	**83%**
Describe a series of pictures depicting an event/action in complete sentences	Picture Description	**83%**	**88%**
Identify the option in which the event is depicted in the correct chronological order	Sequence Identification	**38%**	**36%**
Describe the series of pictures in chronological order	Sequence Identification	**16%**	**15%**

CLASS 1 – HINDI: ITEM SPECIFIC PERFORMANCE (2/3)

Item descriptor	Strand	CONTROL	TREAT-MENT
Identify the alphabet - (High frequency) - र	Letter Reading	**78%**	**94%**
Identify the alphabet - (High frequency) - क	Letter Reading	**86%**	**98%**
Identify the alphabet - (High frequency) - ह	Letter Reading	**70%**	**87%**
Identify the alphabet - (Mid frequency) - त	Letter Reading	**77%**	**93%**
Identify the alphabet - (Mid frequency) - स	Letter Reading	**66%**	**91%**
Identify the alphabet - (Mid frequency) - ल	Letter Reading	**73%**	**96%**
Identify the alphabet - (Low frequency) - फ	Letter Reading	**76%**	**85%**
Identify the alphabet - (Low frequency) - थ	Letter Reading	**59%**	**73%**
Identify the alphabet - (Low frequency) - ज	Letter Reading	**73%**	**91%**
Read 2 letter sight words - (Familiar) - कल	Word Reading	**64%**	**89%**
Read 2 letter sight words - (Familiar) - रख	Word Reading	**56%**	**80%**
Read 3 letter sight words - (Familiar) - समय	Word Reading	**48%**	**74%**
Read 3 letter sight words - (Familiar) - कलम	Word Reading	**51%**	**82%**
Read 2 letter sight words - (Unfamiliar) - यश	Word Reading	**46%**	**70%**
Read 2 letter sight words - (Unfamiliar) - पथ	Word Reading	**50%**	**69%**

CLASS 1 – HINDI: ITEM SPECIFIC PERFORMANCE (3/3)

Item descriptor	Strand	CONTROL	TREAT-MENT
Match the sentence (of 3 words) to correct picture	Picture Matching	**65%**	**83%**
Read a simple sentence (of 4 words)	Word Reading	**38%**	**64%**
Match the sentence (of 4 words) to correct picture	Picture Matching	**60%**	**78%**
Read 1st paragraph (4 lines) from the short narrative text	Reading Comprehension	**27%**	**52%**
Read 2nd paragraph (3 lines) from the short narrative text	Reading Comprehension	**27%**	**51%**
Retrieve directly stated information from the beginning of the text	Reading Comprehension	**37%**	**47%**
Retrieve directly stated information from competing options from middle of the text	Reading Comprehension	**30%**	**43%**
Write 2 letter sight words - (Familiar) - छोटा	Dictation	**36%**	**59%**
Write 2 letter sight words - (Familiar) - बुरा	Dictation	**32%**	**39%**
Write 2 letter sight words - (Familiar) - लाल	Dictation	**52%**	**73%**
Write 3 letter sight words - (Familiar) - किताब	Dictation	**29%**	**37%**
Write 3 letter sight words - (Familiar) - कमरा	Dictation	**43%**	**69%**
Write 4 letter sight words - (Familiar) - रविवार	Dictation	**26%**	**31%**
Sentence Writing – at least ONE complete, relevant without any grammatical mistakes	Picture Description	**22%**	**27%**

GRADE 1 HINDI - SCALE SCORE DISTRIBUTION – TREATMENT SCHOOLS

SCHOOL NAME	SCORE
GPS DHIRPUR	481
GPS SINGHPURA	450
GPS TIKRI	448
GPS SHAHABAD No. 4	440
GPS UMRI	424
GPS THANESAR - 2	422
GPS KISHAN PURA	415
GPS GUMTHALA GARHU	413
GPS NALVI	411
GPS FATUPUR	406
GPS MAJRI KALAN	405
GPS SUDHPUR	401
GPS LANDI	400
GPS BEHOLI	394
GPS PIPLI	376
GPS GMSPS ISMAILABAD	376
GPS SARASWATI KHERA	375
GPS No. 2	371
GPS NAISI	368
GPS BAKHLI	350
GPS GOGPUR	350
GPS TIBBA FARM	347
GPS PALWAL	345
GPS KALYANA	342