LiftEd Transforming Foundational Literacy and Numeracy

EDTECH ACCELERATOR

LiftEd EdTech Accelerator

Impact on Learning Outcomes Study

टॉप पेरेंट
सीखने सिखाने के नए तरीके

Uttar Pradesh | 2026

Principal Investigator
Prof. Tarun Jain
Professor of Economics, Indian Institute of Management Ahmedabad

This study was conducted by Educational Initiatives under the guidance of
Tarun Jain - Professor of Economics and Reserve Bank of India Chair in Finance &
Economics at the Indian Institute of Management Ahmedabad.

Table of Contents

Acknowledgements

We express our profound appreciation to those instrumental in conducting this study and compiling the report. Their support has been indispensable to this endeavour.

We extend our gratitude to the Principal Investigator, Professor Tarun Jain, Associate Professor of Economics and the Reserve Bank of India Chair in Finance and Economics at the Indian Institute of Management Ahmedabad. His guidance in designing the study, overseeing the creation of robust data collection tools and ensuring the quality of insights has been pivotal.

We express our gratitude to the LiftEd EdTech Accelerator consortium partners - Michael & Susan Dell Foundation, Reliance Foundation, UBS Optimus Foundation and British Asian Trust for their flexibility to innovate and generous support throughout the project. We also extend our gratitude to acknowledge the support and counsel from Central Square Foundation.

We extend our deepest gratitude to the team at Educational Initiatives for their expertise in conducting the study, generating insights, and report writing. We are especially grateful to Sridhar Rajagopalan (Co-Founder), Pranav Kothari (CEO), Gayatri Vaidya (Associate Vice President), Arvind Singh (Associate Vice President), and Ei's Product team led by Sangeeta Oak (Associate Vice President) for their leadership and guidance. We would also like to acknowledge Dipak Chalakkal, Vaishali Tiwari, Prashant Kapoor, and Aparna Sanjay, along with the field Operations team and enumerators who conducted the assessments, whose dedication and efforts were instrumental in making this study a reality.

Finally and most importantly, we are indebted to the participants of the study for their time and valuable contributions and the EdTech portfolio partner: Top Parent for their support in facilitating the planning and coordination of the study.

We hope the study findings will inform stakeholders' decisions on designing and implementing effective EdTech interventions and strengthen the collective evidence on EdTech for foundational learning.

In deep gratitude,
Educational Initiatives

Glossary & Abbreviations

Term	Description
Addition/Subtraction Process	Refers to tasks that assess students' ability to solve multi-digit or column-based addition and subtraction problems, requiring correct sequencing of procedural steps such as carrying over or borrowing. These tasks go beyond fact recall and capture procedural fluency in formal arithmetic methods (e.g., solving 47 + 38 using column addition).
Attrition Buffer	Allowance built into the study's sample design to account for the expected loss of participants over time. As the study was longitudinal and conducted over several months, some students present at baseline were unavailable at endline due to absence or mobility and therefore an attrition buffer was included to preserve statistical power.
Conceptual Skills[1]	Refers to competencies that require application, reasoning, integration of multiple concepts and problem solving. These skills typically involve multi-step, non-routine, or contextualised tasks and may include spatial, or language-heavy demands, such as solving word problems or applying arithmetic concepts to real-life situations. The tasks from the Numeracy tool classified as conceptual Skills for the purposes of this study are Addition Process (2 digit, column addition), Subtraction Process (2 digit, column addition), Counting in Bundles, Missing Number, Word Problems and Shape Recognition.
Confidence	Refers to the degree of certainty that observed results reflect a true underlying effect rather than chance. It is influenced by sample size, outcome variability, study design and precision of estimates, often reflected through p-values and confidence intervals.
Confidence Intervals (CI)	Represent a statistical range around an estimated effect that conveys the degree of uncertainty in the estimate. They indicate the range within which the true population effect is likely to lie, assuming the underlying model is correct (e.g., an average of estimate of 0.7 year with a 95% confidence interval of 0.6 to 0.9).
Difference-in-Differences (DiD)	Analytical method used to estimate programme impact by comparing changes in outcomes over time between Intervention and Comparison groups. This approach helps account for baseline differences and common time trends affecting both groups.
Intent-to-Treat (ITT)	Impact estimate that captures the effect of being assigned to the intervention, irrespective of actual participation or usage levels. ITT estimates preserve the original group assignment and provide a conservative measure of programme effectiveness under real-world conditions.
Intra-Class Correlation (ICC)	Refers to the degree to which students within the same class or school resemble one another in terms of learning outcomes, reducing the amount of independent information contributed by each student.

1 Ministry of Education, Government of India. (2021). National initiative for proficiency in reading with understanding and numeracy (NIPUN Bharat): Mission guidelines. https://www.education.gov.in/sites/upload_files/mhrd/files/nipun_bharat_guidelines.pdf .

Term	Description
Minimum Detectable Effect Size	The smallest difference between Intervention and Comparison groups that the study is statistically powered to detect, given assumptions about sample size, outcome variability, clustering and significance levels.
Outcome Variability/ Variance	Refers to the amount of natural variation in outcomes across individuals rather than uniform responses. Higher variability in student performance increases uncertainty in estimates and typically requires larger samples to detect programme effects.
Power	Probability that a study will correctly detect a true effect of a specified size if it exists in the population. Studies are commonly designed with power levels of 80 percent or higher to reduce the likelihood of false negatives.
Procedural Skills2	Refers to tasks that primarily assess recall, recognition and the execution of well-rehearsed procedures, with minimal reasoning demands. These tasks typically focus on basic number sense and fact recall, involve limited interpretation or transfer and are relatively language-light. For the purposes of this study, the following tasks in the numeracy tool are classified as Procedural Skills: Number Comparison, Addition Facts (untimed), Addition Facts (timed), Subtraction Facts (untimed) and Subtraction Facts (timed).
Quasi-Experimental Design (QED)	Study design used to estimate the effects of an intervention when random assignment of participants to treatment and comparison groups is not feasible. Instead, the design relies on non-randomly formed groups and applies statistical or design-based techniques (such as matching or difference-in-differences) to control for pre-existing differences and approximate causal inference.
Standard Deviation (SD)	Statistical measure of variability that indicates how spread out individual observations are around the mean. When used to express effect sizes, SD units standardise differences across groups or time points, allowing impacts to be compared across outcomes measured on different scales.
Treatment-on-the-Treated (ToT)	Impact estimate that reflects the effect of the intervention on participants who actually received or engaged with it as intended. ToT estimates adjust for differential usage or compliance and are typically derived using assignment to treatment as an instrument for actual participation.

2 Ministry of Education, Government of India. (2021). National initiative for proficiency in reading with understanding and numeracy (NIPUN Bharat): Mission guidelines. https://www.education.gov.in/sites/upload_files/mhrd/files/nipun_bharat_guidelines.pdf .

1. Executive Summary

This chapter outlines the purpose, approach and key findings from the independent evaluation of the Top Parent's EdTech solution in Uttar Pradesh, examining its effectiveness in improving foundational numeracy outcomes.

1.1 About the Study

Foundational Literacy and Numeracy (FLN) underpin lifelong learning in India, yet significant learning gaps emerge early and persist despite recent strides under NEP 2020, NIPUN Bharat Mission and increased FLN investments. Inclusive EdTech presents a strong opportunity to strengthen FLN at home and in schools, supported by growing smartphone access. However, most existing solutions remain poorly aligned with the languages, needs and contexts of low-income communities.

The LiftEd EdTech Accelerator was created to bridge this gap by supporting contextually relevant, pedagogically sound EdTech for "Low-Income Bharat." Working with governments and ecosystem partners, it enables large-scale adoption through funding, mentorship and technical assistance for eight leading EdTech organisations. A robust evaluation agenda focused on learning outcomes, impact and user experience has generated actionable insights. Collectively, LiftEd-supported solutions have reached over 5.5 million users, directly benefiting more than 3 million children in improving foundational learning.

To build evidence on EdTech's effectiveness in FLN, the Accelerator commissioned an independent evaluation to identify what works, its impact on learning outcomes and key enablers for scale. Conducted by Educational Initiatives (Ei) with Central Square Foundation, the study assessed selected solutions' impact on early mathematics across grades and geographies. As the outcome evaluator, Ei consolidates findings at the partner level, measuring the impact on learning outcomes as a result of the solution usage.

The **EdTech solution evaluated in Uttar Pradesh is Top Parent** – a free, lightweight B2C learning app that helps low-income families build early literacy and numeracy through parent engagement, personalized content and continuous assessments. It offers age-mapped videos, gamified worksheets, assessments and vernacular, audio-supported access to help children learn anytime with minimal data and connectivity.

1.2 Study Design

The study assessed the impact of **Top Parent's at-home FLN intervention** on numeracy outcomes for Grades 2 and 3 students in Ambedkar Nagar, Ayodhya, Bahraich and Sultanpur districts in Uttar Pradesh. It used a **longitudinal quasi-experimental design with a difference-in-differences approach**[3] to compare learning growth between intervention and comparison groups over time, with the resulting impact estimates subsequently converted into Equivalent Years of Schooling (EYOS) to aid interpretation.

3 The DiD effect size is calculated as: [Avg Delta_Intervention (Δi) - Avg. Delta_Comparison (Δc)] / Pooled Standard Deviation]

The Baseline round was conducted in September - October 2024 and Endline round in November - January 2025. Assessments were administered through the Ei NEEV application, capturing student responses via audio and touch-based inputs.

1.3 Key Findings

1.3.1 The Top Parent intervention delivered consistent gains in foundational numeracy, with limited shifts in conceptual skills

Students demonstrated improvements in procedural tasks, with effect sizes clustering between 0.10–0.13 SD. The strongest gains were observed in Number Comparison (0.13 SD) and Subtraction Facts (0.13 SD), signaling strengthened number sense and arithmetic fluency. These results indicate that Top Parent effectively reinforced early numeracy building blocks such as quantity comparison, fact retrieval and basic operational fluency. In contrast, conceptual competencies - including Word Problems and Shape Recognition - showed non-significant gains.

The EYOS analysis indicates meaningful learning acceleration of ~1.4 times relative to Business-As-Usual instruction. The intervention generated learning gains equivalent to a substantial fraction of a year of schooling over the study period, demonstrating that students progressed faster than peers in comparison schools. These findings suggest that the model not only improves absolute learning levels but also meaningfully compresses the time required to achieve expected grade-level competencies, highlighting its potential to address foundational learning gaps in low-performing contexts.

1.3.2 At a grade level, patterns show steady gains for Grade 2 in number sense and arithmetic skills, while Grade 3 students recorded more uneven progress

Grade 2 learners exhibited the strongest improvement across skills, with statistically significant gains in Number Comparison (0.16 SD), Subtraction Facts (0.16 SD) and Subtraction Process (0.16 SD), alongside positive movement in most remaining tasks. These results suggest that younger students were more responsive to structured digital practice, particularly in foundational and procedural domains. Grade 3 outcomes were more variable: while gains were observed in Missing Number (0.10 SD) and timed Subtraction Facts (0.14 SD), improvements were smaller in magnitude. This pattern indicates that higher grade learners may require more targeted or intensive support to translate practice into broader conceptual growth[4].

4 Powell, S. R., & Fuchs, L. S. (2015). Intensive intervention in mathematics. ERIC. (ED561207); Svane, R. P., Willemsen, M. M., Bleses, D., Krøjgaard, P., Verner, M., & Nielsen, H. S. (2023). A systematic literature review of math interventions across educational settings from early childhood education to high school. Frontiers in Education, 8, 1229849. https://doi.org/10.3389/feduc.2023.1229849. As students progress to higher grades, the average impact of mathematics interventions tends to diminish, and learners with persistent difficulties increasingly require more intensive and individualized instructional approaches—beyond general practice-based exposure—to achieve substantive, conceptually oriented learning gains.

1.3.3 Boys showed stronger gains in fluency tasks, while girls demonstrated steadier improvements in number sense and counting

Gender-disaggregated findings reveal distinct performance patterns. Boys recorded larger gains in arithmetic fluency—particularly in Addition Facts (0.23 SD untimed; 0.17 SD timed) and Subtraction Facts (0.15–0.16 SD). However, their performance in conceptual tasks, including Word Problems (–0.04 SD) and Shape Recognition (0.01 SD), remained flat or negative. Girls, in contrast, displayed more balanced improvements, with notable gains in Number Comparison (0.23 SD) and Counting in Bundles (0.17 SD), alongside steady progress in Missing Number (0.09 SD) and process-based tasks (0.08 SD each in Addition and Subtraction). While girls showed smaller jumps in fluency, their progress was more uniform across domains.

1.3.4 Gains in procedural skills outpaced conceptual skills, with limited mobility among lowest performers

Movement across proficiency bands reinforces the effect-size findings. In Number Comparison, only 17% of Intervention L0[5] learners remained at L0 (from Baseline to Endline) compared to 20% in the Comparison group, while 67% reached L4 versus 52% in Comparison. Similar trends appear in fluency tasks such as Addition Facts, where 60% of Intervention L0 students moved to L4 compared with 49% in Comparison. Mid-level performers (L2–L3) in the Intervention group also showed markedly higher upward movement than their Comparison counterparts. In conceptual tasks, mobility was more limited for both groups, but Intervention students still showed an advantage - for example, fewer L0 students remained stagnant in Word Problems (23% vs. 36%). Overall, gains were stronger in procedural tasks than in conceptual tasks. Nevertheless, the intervention consistently supported upward movement among low- and mid-performing learners, even in domains where overall mobility remained constrained.

The Top Parent intervention shows clear potential for strengthening early numeracy, particularly number sense, arithmetic fluency, and foundational processes among younger learners, regular users, and those starting with lower proficiency. Upward movement of Intervention students point to the value of structured, home-based digital practice in supporting core skill growth. At the same time, lower gains in conceptual tasks and variations across grades and genders suggest opportunities for deeper conceptual scaffolding and more tailored support for older learners and girls' fluency development.

5 Categories L0–L4 classify student performance from zero scores to highest proficiency, based on accuracy percentages and fluency relative to the baseline average. These levels help track patterns of improvement or decline across the evaluation timeline. L0 = Zero scorers, L1 = 0 to 25% (accuracy tasks) and 0 to 0.25x Baseline Avg (fluency tasks). Further details on the classification are presented in Section 4.4 of this report.

2. Background of the Study

2.1 Inclusive EdTech for Foundational Learning

Foundational Literacy and Numeracy (FLN)—the ability to read, write, comprehend and perform basic mathematical operations by the end of Grade 3 - is widely recognised as the cornerstone of lifelong learning.[6, 7, 8, 9] In India, significant learning deficits emerge early, particularly among children aged 4 to 8, with large proportions of students struggling to acquire basic skills. Repeated rounds of the Annual Status of Education Report (ASER) over the past decade have consistently highlighted low levels of foundational learning. These early gaps make it increasingly difficult for children to grasp more complex concepts as they move through higher grades. The most recent ASER data, however, points to green shoots of progress - improvements in basic reading levels for Grade 3 government students being the highest since 2005 with basic arithmetic levels also showing substantial improvement.[10] Yet, the depth and persistence of learning deficits across the system make clear that far more remains to be done.

To address these challenges, the Government of India has launched multiple initiatives to strengthen FLN, including NEP 2020, which frames FLN as a five-year continuum and emphasises curriculum reform, technology integration and teacher capacity-building. NIPUN Bharat (2021) sets FLN targets for preschool to Grade 3 by 2026–27, supported by initiatives such as the CBSE Reading Mission, the National Curriculum Framework for the Foundational Stage and Jaadui Pitaara, which together provide age-appropriate, multilingual learning resources. Progress is monitored through large-scale assessments like NAS and FLS, while increased funding under Samagra Shiksha and state-level programmes, often implemented with private and civil society partners, aim to further bridge learning gaps[11, 12] Further, budget allocations for education under Samagra Shiksha have increased, from ₹31,050 crores (~4200 million USD) in FY 2021-22 to ₹41,250 crore in FY 2025-26.[13] Concurrent to the central government initiatives, state governments, in collaboration with private partners and civil society organisations, have implemented tailored initiatives to bridge learning gaps.

In this context, educational technology (EdTech) has emerged as a potential multiplier, harnessing digital tools to support teaching and learning both in classrooms and at home. A recurring question, however, is whether EdTech is truly accessible to all, especially children from low-income households. Emerging data suggests that access to digital

---

6 Ministry of Human Resource Development, Government of India. (2020). *National Education Policy 2020.* https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
7 Institute for Competitiveness. (n.d.). *State of Foundational Literacy and Numeracy in India.* https://www.competitiveness.in/wp-content/uploads/2021/12/Report_on_state_of_foundational_learning_and_numeracy_web_version.pdf
8 Ministry of Education, Government of India. (n.d.). *About Foundation Literacy and Numeracy.* https://diksha.gov.in/fln.html.
9 Sinha, A (2023). Maximising India's demographic dividend through foundational literacy and numeracy. *Hindustan Times.*(https://www.hindustantimes.com/ht-insight/knowledge/maximising-india-s-demographic-dividend-through-foundational-literacy-and-numeracy-101699332886168.html).
10 Annual Status of Education Report (ASER). (2024). *ASER 2024 National Findings* https://asercentre.org/wp-content/uploads/2022/12/ASER-2024-National-findings.pdf
11 Ministry of Human Resource Development, Government of India. (2020). *National Education Policy 2020.* https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
12 Storyweaver (n.d.). *CBSE Reading Mission.* https://storyweaver.org.in/en/about/campaigns/cbse-reading-mission
13 CNBC TV 18. (2025). *Budget 2025: National Education Mission receives outlay of ₹41,250 crore* https://www.cnbctv18.com/budget/budget-2025-national-education-mission-samagra-shiksha-abhiyaan-receives-outlay-of-rs-41250-crore-19548980.htm

infrastructure in these contexts is becoming increasingly feasible, particularly through smartphones. Studies indicate that **90% of households** have access to at least one smartphone and in **75% of these households**, children regularly use the device - typically spending around an hour on it each day.[14] This rise in smartphone penetration, coupled with growing internet availability, presents a significant opportunity to leverage EdTech solutions to improve learning outcomes and narrow educational gaps.

Building on this opportunity, EdTech offers innovative solutions to address critical educational challenges, such as varying teacher quality, diverse learning levels within classrooms and limited access to quality instructional resources. By equipping teachers with tools for effective pedagogy and enabling parents to support their children through interactive content and progress tracking, well-designed, pedagogically sound EdTech solutions have the potential to significantly improve learning outcomes.

Globally, there is emerging evidence on the potential of EdTech to support learning at home. The Global Learning XPRIZE Competition, launched in 2014, incentivised teams from around the world to create open-sourced, scalable software that empowers children to achieve foundational learning skills and saw learning gains for both literacy and numeracy across competing solutions.[15] Similarly, Angrist, Bergman and Matsheng provide experimental evidence on strategies to support learning when schools close.[16] Using a randomised control design, they tested two low-technology interventions in Botswana – SMS messages and phone calls – with parents to support their child's learning and found that combined treatment improves learning by 0.12 standard deviations. This translates to 0.89 standard deviations of learning per USD 100, ranking among the most cost-effective interventions to improve learning.

Despite this promise, most current EdTech solutions in India are designed primarily for middle- and high-income users. Content is often in English, misaligned with the lived realities and languages of children from low-income communities and offered at price points that place it beyond their reach. As a result, the children who could benefit most from effective EdTech are often the least likely to access it. Unlocking EdTech's transformative potential for "Low Income Bharat" therefore requires inclusive, affordable solutions that are contextually relevant, explicitly address foundational learning and are backed by rigorous evidence of effectiveness. Bridging these gaps will ensure that EdTech becomes a critical lever for equitable and impactful education in India. It was with this objective that the **LiftEd EdTech Accelerator** was set up.

2.2 LiftEd EdTech Accelerator

To bridge the gaps as defined above and to leverage the opportunity that India has, a consortium of non-profit and philanthropic organisations have set up a LiftEd EdTech Accelerator, a two-year initiative from April 2023-25, to support foundational learning of children using EdTech. The Accelerator was set up to support the NIPUN Bharat mission to significantly shape the future of tech-based learning at home for foundational literacy and numeracy in India by reaching 2.5 million children by 2025.

14 Central Square Foundation. (2026). *Bharat Survey for EdTech (BaSE) Report 2026.* Bharat Survey for EdTech (BaSE) Report 2025

15 Global Learning X Prize. (n.d.). *Global Learning X Prize: Executive Summary* https://assets-us-01.kc-usercontent.com/5cb25086-82d2-4c89-94f0-8450813a0fd3/fc467c7f-d8bd-4d05-bba3-2aa9b06833fb/GLEXP_Executive%20Summary.pdf

16 Angrist, N., Bergman, P. & Matsheng, M. (2022). Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour*, 6, 941-950. https://doi.org/10.1038/s41562-022-01381-z

The LiftEd EdTech Accelerator is anchored by Michael & Susan Dell Foundation, Reliance Foundation and UBS Optimus Foundation as Founding Partners, the British Asian Trust as the Programme Leader and Central Square Foundation as the Design and Technical Partner.

The Accelerator aimed to catalyse the supply of contextually relevant and pedagogically sound learning solutions, generate compelling evidence on their efficacy, work with governments to enhance the efficacy of EdTech adoption and create public goods to address systemic challenges in the ecosystem.

The Accelerator aimed to support eight high-quality EdTech solutions for two years through impact-focused grant funding, dedicated mentorship and capacity-building support to unlock the full potential of the EdTech solutions. The solutions were onboarded into three cohorts, each addressing key challenges in the Indian EdTech ecosystem. The cohorts focused on:

1. Scale – products looking to discover and unlock new pathways to scale - ThinkZone
2. Engagement – products seeking strategies to deepen engagement with the users - Chimple, Ei Mindspark, Pratham, Rocket Learning, Top Parent
3. Product Contextualisation – products developing pedagogically sound and contextually relevant solutions specifically for low-income India - Amira Learning and Sesame Workshop India (SWI)

On the demand side, the Accelerator focused on driving the adoption and institutionalisation of tech based home learning for FLN within State Governments, while also exploring innovative pathways for EdTech integration through partnerships with retail channels, such as gig economy organisations and self-help groups (SHGs).

To tackle the challenge of limited existing evidence on 'what works' in EdTech and to allow for ongoing innovation and progress, the Accelerator's evidence generation agenda included

1. Learning Outcomes Evaluation - to assess the impact on student learning outcomes for ThinkZone, Top Parent and Ei Mindspark.
2. Impact of Acceleration Study - to capture the effectiveness of the strategies implemented within the Accelerator (published)
3. Insights on User Experience Study - a qualitative analysis that gathers feedback from end users on key aspects of the EdTech programme lifecycle, including acquisition, onboarding, engagement and retention.

These evaluations have been conducted under the supervision of the Principal Investigator, Prof. Tarun Jain (Reserve Bank of India Chair Professor of Economics at Indian Institute of Management, Ahmedabad) by experts from Sambodhi Research (qualitative study) and Educational Initiatives (quantitative study) to provide actionable insights to inform future interventions and improvements.

Over 2023-2025, the LiftEd EdTech Accelerator has collaborated with eight leading EdTech organisations to advance the future of tech-enabled at-home learning for foundational literacy and numeracy (FLN) in India. Collectively, these solutions have reached over **5.5 million** users, with more than **3 million** directly benefiting from features and innovations developed through the Accelerator's support.

This report provides insights from the **Learning Outcomes Evaluation Study** for **Top Parent**.

2.3 Overview of EdTech Partner Model

Research[17] demonstrates that parental engagement is critical to a child's learning journey and has been associated with improved school performance and greater social and emotional development. However, parental engagement in low-income communities is often a challenge given low textual literacy and availability of resources. The driving principle of the Top Parent app is to actuate parental engagement – the belief is that if parents are provided with necessary resources and skills, they can use digital technology to engage with and support their children's education more meaningfully. The focus is on parents of children aged 3-8 years in low-income households (monthly incomes between ₹ 10,000-30,000). Currently, the app is available in Hindi and Marathi and covers the subjects of English, Hindi, Marathi, GK and Mathematics.

The Top Parent app is free to use and utilises a user-centric design to support both parents and children. Frequent user studies and dipstick surveys are conducted to ascertain what the user would find useful and insights from this are funnelled into app development and roll-out. The app was developed in 2020, when schools were closed due to the pandemic, and it was critical to build a programme to support parents. The app is designed as a progressive web app (PWA) and an Android app, which ensures that parents with basic smartphones having limited storage capacity (as is often the case in low-income households) can access the app. The app has two sections – for parents and children. Parent facing content ensures that parents are empowered with knowledge, skills and learning content to support child learning and well-being at home through information on foundational learning, socio-emotional learning, nutrition, child protection and child health. For children, multimedia content in the form of quizzes, videos, gamified worksheets and assessments are provided. Concepts are mapped to competencies in numeracy and literacy as per the NCERT curricula.

Top Parent is a B2C solution – i.e., parents come to know about the app through online platforms (YouTube videos, Google ads) and social networks (neighbours, friends, family). Parents then download the app on their smartphones. A computerised pre-test mapped to age and competency is provided to the child. Based on this pre-test, the interactive content is unlocked, which helps to introduce, reinforce and scaffold learning. In addition, continuous assessment tracks performance of the child. A weekly report card is generated, which communicates to the parent via WhatsApp in simple and visually engaging (e.g., colours, happy/sad face) on how the child is doing. This is because often parents in lower income households have low textual (and digital) literacy. The weekly report card is accompanied by a very specific nudge on whatsapp telling the parent what can be done to reinforce/scaffold and to sustain engagement. An AI chatbot on WhatsApp responds to queries of parents.

17 Lin, Q (2003). Lin, Parent Involvement and Early Literacy / Family Involvement Research Digests / Publications Series / Publications & Resources / HFRP - Harvard Family Research Project. (C) Presidents and Fellows of Harvard College. https://archive.globalfrp.org/publications-resources/publications-series/family-involvement-research-digests/parent-involvement-and-early-literacy.

Under the LiftEd Accelerator, the focus is on introducing and testing the features of this AI chatbot and live streaming for blended learning and spoken English classes to improve engagement on the product. The accelerator is being rolled out pan-India, with a focus on Hindi and Marathi speaking belts.

Product overview

- **Subject areas:** Math and Language
- **Languages:** Hindi and Marathi
- **Grades**: 1, 2, and 3

Theory of Change

1. Include NCERT mapped multimedia resources for children and videos for parents

- Multimedia content in the form of quizzes, videos and gamified worksheets to engage children and help them learn
- Parent-facing content in the form of videos to help support their child

2. Grounded in accelerated learning framework pedagogy

- Computerised pre-test mapped to age and competency
- Live streaming classes support and scaffold learning
- Continuous assessment tracks performance and generates report cards

3. Drive engagement through customised nudges and gamification

- Drive and sustain engagement basis real time data fed to generate customised nudges for parents
- Keeps parents engaged through gamified quizzes, rewards and leaderboards
- AI chatbot integrated with Jugalbandi to respond to users

Figure 1: Top Parent Model

Despite challenges common to B2C EdTech - fragmented audiences, high churn and resource constraints - Top Parent's human-centered design, strong behavioural nudging architecture and evidence-driven development approach position it as a promising solution for improving early learning at scale.

3. Evaluation Design and Approach

This section outlines the study's research design, detailing how the quasi-experimental setup, sampling approach and assessment methods were structured to measure Top Parent's impact on numeracy outcomes. It also summarises the timelines, data collection processes and safeguards used to ensure validity and reliability of findings.

3.1 Study Design

3.1.1 Overview

The study employed a quasi-experimental design (QED), with an Intervention group and a Comparison group identified to benchmark impact. This study seeks to answer the question:

> What impact does the use of the intervention programme have on student learning outcomes?

The Intervention group included students who received access to the Top Parent programme. The Comparison group comprised similar students who did not receive the programme during the study period. This setup enabled a clear assessment of the effect of EdTech exposure by comparing progress between the two groups with comparable baseline characteristics. To estimate causal impact, the study used a **difference-in-differences (DiD) approach**, measuring changes in learning outcomes over time while controlling for time-invariant factors.

Students from both Grades 2 and 3 were included in the study, although sampling was not conducted at the grade level. Assessments were carried out in two rounds to measure learning outcomes at the start (baseline) and end (endline) of the evaluation period.

3.1.2 Timelines

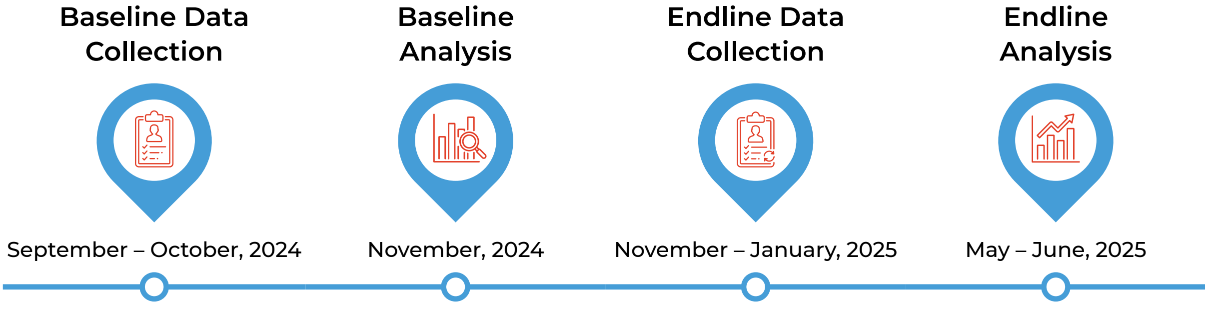

Figure 2: Study Timelines

3.2 Sampling Strategy

3.2.1 Sample Design

The evaluation was conducted in Ambedkar Nagar, Ayodhya, Bahraich and Sultanpur districts. Given that Top Parent's users are acquired through online ads, their initial user base was highly geographically dispersed. In order to create a concentrated sample for the Accelerator evaluation, the team engaged Awadh Youth Collective to enroll students from select blocks in Ambedkar Nagar, Ayodhya, Bahraich and Sultanpur districts in Uttar Pradesh.

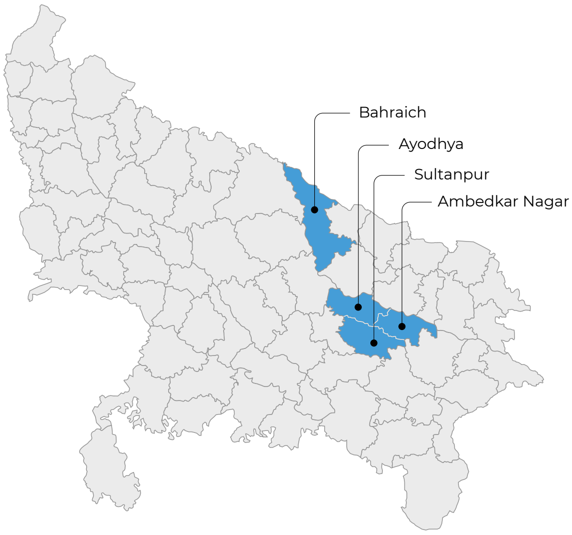

Figure 3: Geography covered

The partner was initially tasked with enrolling 6,000 Grade 2 and 3 students (approximately 1,500 per district), based on an assumed average cluster size of 20 students per community. As recruitment progressed, it became evident that the feasible enrolment per community was closer to 10 students. In response, the overall target was revised downward to 4,400 students. The partner successfully enrolled 4,607 students across more than 900 communities.

The following considerations were used for sample size calculation:

Note: Factors marked with*apply only to the Intervention group.

Figure 4: Sample Size Parameters

Based on the above factors, the final sample size for both the Intervention and Comparison groups was determined to be **3,854 students from at least 385 Intervention schools and 1,990 students from 199 Comparison schools**. Details regarding the sampling process are presented in Annexure 6.1 to 6.3.

3.2.2 Selection of Intervention and Comparison Groups

Top Parent recruited students for its EdTech Accelerator intervention through the Awadh Youth Collective (AYC), a network of community organisations, enrolling children from selected blocks across four districts - Akbarpur and Jalalpur (Ambedkar Nagar); Ayodhya Town, Masodha and Milkipur (Ayodhya); Kaiser Ganj, Fhakharpur and Risia (Bahraich);

and Dubepur and Sultanpur City (Sultanpur). AYC field partners conducted door-to-door mobilisation to onboard Grade 2 and 3 students from local communities.

Based on enrolment patterns observed in previous FLN assessments in Uttar Pradesh, AYC-enrolled students were expected to be concentrated in a limited number of nearby schools. These schools were therefore treated as proxy intervention schools. Comparison schools were drawn from other blocks within the same districts, specifically, schools with no AYC-enrolled students but with similar socio-economic and demographic characteristics, to minimise the risk of spillover. Coarsened Exact Matching (CEM) was subsequently applied to identify a set of comparable schools for the comparison sampling frame.

As the Intervention and Comparison groups were imbalanced in size, a combination of many-to-one and one-to-many matching was employed to maximise overlap, maintain operational feasibility and ensure cluster sizes remained below 10 students per school. This resulted in a many-to-many matching structure. Matching could be implemented for only 341 schools due to data gaps in the community partner-provided records and the presence of students attending unregistered or non-UDISE schools. However, as all sampled students were drawn from a fixed set of AYC communities, the matched schools are considered to be broadly representative of the intervention-school pool relevant for this sampling exercise. Further details are provided in Annexure 6.2.

3.2.3 Final Sample Achieved

Since the study was longitudinal, tracking the performance of the same students over the evaluation period, students were excluded from the final dataset if they fell into one of the following categories at the end of data collection:

- Attempting only one subject in a given round due to lack of consent, operational challenges, or other constraints
- Absent or unsynced audio response recordings / assets
- Presence in the Baseline but absence in the Endline

The final sample is as follows:

Table 1: Details of Final Sample

Arm	Sample Size	Minimum Required	Final Sample Achieved
Intervention	3854	304	1891
Comparison	1990	304	1126

Since the minimum required sample size was achieved, no changes were needed to the MDES of the study. A detailed summary of the sampling process is provided in Annexure 6.3.

3.3 Data Collection Methods and Tools

This section outlines the assessment tools, data collection processes and safeguards employed to ensure accuracy and reliability of findings

3.3.1 Assessment Tool

A contextualised version of the **Early Grade Mathematics Assessment (EGMA)** was used for this study. The same assessment tool was administered to students in Grades 2 and 3 and remained unchanged across both the Baseline and Endline rounds. This was feasible given the EGMA-based FLN tool assesses largely common concepts and skills across Grades 2 and 3, with the exception of multiplication and division facts, which are typically introduced in Grade 3. The trade-off of excluding these Grade 3–specific skills is outweighed by the advantages of using a common tool, which enables the calculation of consolidated scores and effect sizes across both grades. This approach allowed the study to examine the impact of the EdTech solution across two grade levels while expanding coverage without increasing sample size or implementation costs. To ensure contextual relevance, the assessment tools were adapted, contextualised and translated into Hindi.

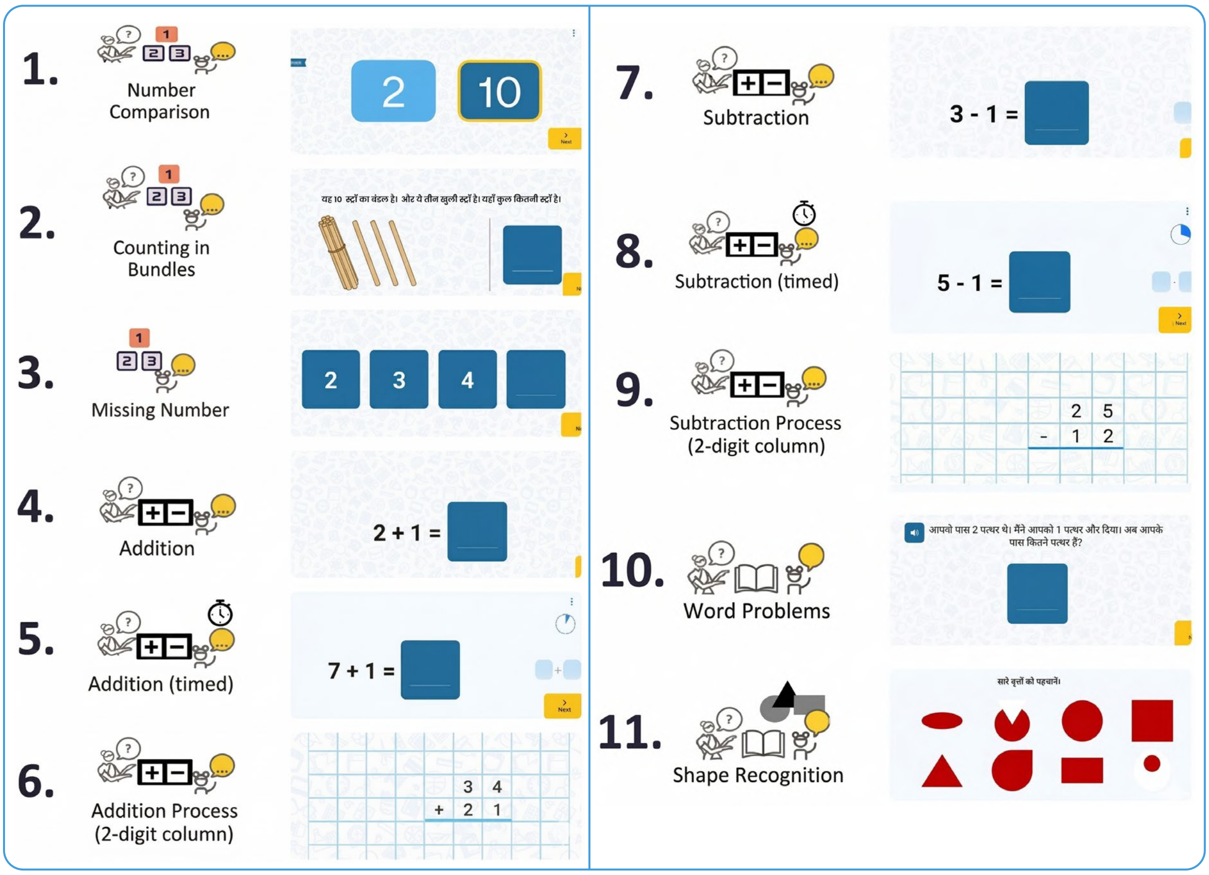

Figure 5: Tasks Assessed

The scaffolded assessments capture foundational skills, ensuring robust measurement of learning outcomes. Further details on the tool are provided in Annexure 6.4.

3.3.2 Data Collection Process

Given that the Top Parent primarily functions as an at-home intervention, data collection was conducted through a community-based model. While each assessment was one-on-one in nature - with students responding independently on their own devices - each enumerator simultaneously supervised two children at a time to ensure smooth administration and adherence to protocols. Dedicated enumerator training sessions were held before each round of assessment.

Assessments were administered using the Ei NEEV application, installed on tablets provided to students at the start of the session. The application included verbal instructions to guide students through the process. Data was captured as touch-based digital entries, which were auto-scored by the application.

3.3.3 Regional Considerations

The operations plan was designed to account for region-specific factors that could affect student availability and participation. Scheduling of assessments took into consideration public holidays, the local harvest season and the school academic calendar. These adjustments ensured that data collection was minimally disruptive, logistically feasible and allowed for smooth test administration across all sampled schools.

3.3.4 Safeguards for Data Quality

Data collected via the Ei NEEV application was stored in a central database and subsequently transferred to a secure evaluation portal. Tasks were processed using pre-validated scripts that had undergone Quality Analysis testing. For timed tasks, items not reached by students were automatically scored as zero to maintain consistency. Once both Baseline and Endline scoring was completed, further verification checks were conducted, including:

- Analysing scores across competencies[18]
- Checking progression patterns[19]
- Identifying anomalies such as outlier distributions[20]

3.4 Challenges and Implications

The study provides relevant evidence on the intervention's impact, though certain limitations mentioned below should be kept in mind when interpreting the findings.

3.4.1 Technical Challenges

- As assessments were conducted in community-based settings, the testing environment was not always fully controlled. Ambient noise from surrounding localities - such as nearby homes and, in a limited number of cases, close proximity of siblings or neighbours also attempting the assessment, occasionally affected the quietness of the assessment setting. However, verification of the audio files confirmed that the background noise had no impact on validity of the assessments.

18 Aggregate and task-level scores were reviewed across competency domains to ensure internal consistency and expected variation in performance.
19 Baseline to Endline score changes were examined to verify plausible learning trajectories at the student and group levels.
20 Score distributions were screened to flag extreme values or irregular patterns that could indicate data or scoring issues.

- Additionally, during the Baseline, technical issues related to tablet–database synchronization also resulted in the loss of certain audio files. Although recovery efforts were made through system backups, a small portion could not be retrieved. By the Endline, these issues were largely resolved for the purposes of this programme , significantly reducing data loss. Given the large sample size achieved in the study, the minimum required sample size was comfortably exceeded, ensuring no risk to the validity of findings. A detailed breakdown of the final sample determination is included in Annexure 6.3.

3.4.2 Sample Attrition and Participation Gaps

Between the Baseline and Endline rounds, a proportion of students attritioned due to migration, or unavailability during the data collection window. These cases were excluded from the final sample, reducing the overall sample size available for analysis[21].

Lee Bounds analysis was conducted to assess the potential impact of student attrition between baseline and endline. As part of this analysis, students who were present at baseline but missing at endline were assigned a score of zero, thereby generating more conservative estimates of the intervention's impact and ensuring that the reported effects are robust to potential bias arising from differential dropout. Lee Bounds estimates[22] are lower than the DiD effects and often reverse sign, with several remaining statistically significant in the negative direction. This suggests the positive impacts are not robust to potential attrition bias and should be interpreted with caution.

3.4.3 Operational Challenges

Accurate identification of students across the Baseline and Endline rounds posed an operational challenge. As assessments were conducted in community based settings, student identification relied on a combination of identifiers, including enrolled school, to determine intervention status. In several cases, enumerators were unable to confidently verify a student's existing details at Endline, requiring the creation of new student IDs in the field. This limited direct ID-level matching between rounds. To address this, an extensive post-collection matching exercise was undertaken using multiple attributes like student name (accounting for spelling variations), age, grade, school and district to reliably link Baseline and Endline records. This process was implemented to minimise misclassification and ensure continuity of student-level analysis across assessment rounds. Where students could not be matched with a sufficient degree of certainty, their records were excluded from the final analytical sample.

21 Demographic Analysis of the students who dropped out of the study after the Baseline is presented in Annexure 6.5.
22 Lee bounds provide a conservative range of treatment effects by adjusting for potential bias from differential attrition, effectively assuming that all individuals who dropped out would have had the lowest possible outcome (e.g., scored zero).

4. Findings

This section presents the key findings from the evaluation, detailing the overall impact of the Top Parent programme on and how these effects vary by grade and gender. It also examines score movements to provide deeper insight into where the intervention drove progress and where challenges persist.

4.1 Overall Impact

The Top Parent programme's numeracy evaluation shows **consistent gains across procedural skills**, with effect sizes clustering in the range of 0.10–0.13 SD. The strongest improvements appeared in **Number Comparison** (0.13 SD) and **Subtraction Facts** (untimed – 0.13 SD; timed – 0.13 SD), suggesting strengthened number sense and arithmetic fluency. Since quantity comparison and fact retrieval are foundational skills that scaffold later mathematical development, these gains indicate that the programme was effective[23] in reinforcing early numeracy building blocks. This pattern also signals that the **intervention is succeeding where home-based practice models** typically have the most leverage—on fluency-oriented competencies that **respond quickly to repeated exposure**.[24]

Addition and Subtraction Processes[25] (0.10–0.12 SD) also showed positive improvements. These tasks require applying operations using multi-step reasoning - such as understanding place value or carrying/borrowing - which indicates some progress in conceptual fluency. However, the observed gains are smaller than those in simple number sense and fact-based tasks. **Counting in Bundles (0.11 SD) and Missing Number (0.12 SD)** – skills that tap into place-value reasoning and number structure – also showed gains, which in turn reinforce the progress in arithmetic processes.

23 Detailed analysis has been shared in Annexure 6.6. Sub-task results are denoted by significance levels as follows: * indicates $p < 0.01$; ** indicates $0.01 \le p < 0.05$; and *** indicates $0.05 \le p < 0.10$.

24 Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Bryant, J. D., & Hoard, M. K. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology, 105*(1), 58–77. https://doi.org/10.1037/a0030127. Experimental evidence from first grade shows that brief, targeted number knowledge tutoring with deliberate, structured practice leads to rapid improvements in basic arithmetic fluency and number skills, indicating that fluency-oriented competencies are particularly responsive to repeated exposure in early elementary contexts.

25 Addition and Subtraction Process (untimed) denotes column addition/subtraction involving one- and two-digit numbers.

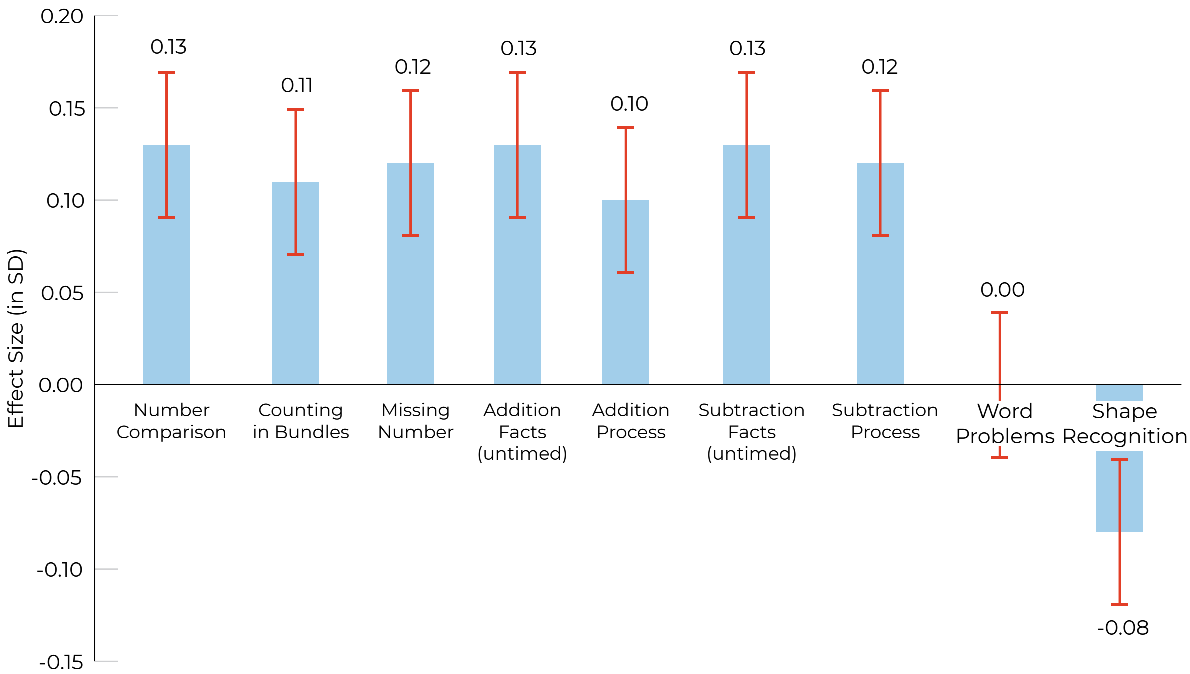

Figure 6: Task wise DiD Effect Sizes (in SD) – Untimed Tasks.

The red lines denote the standard errors.

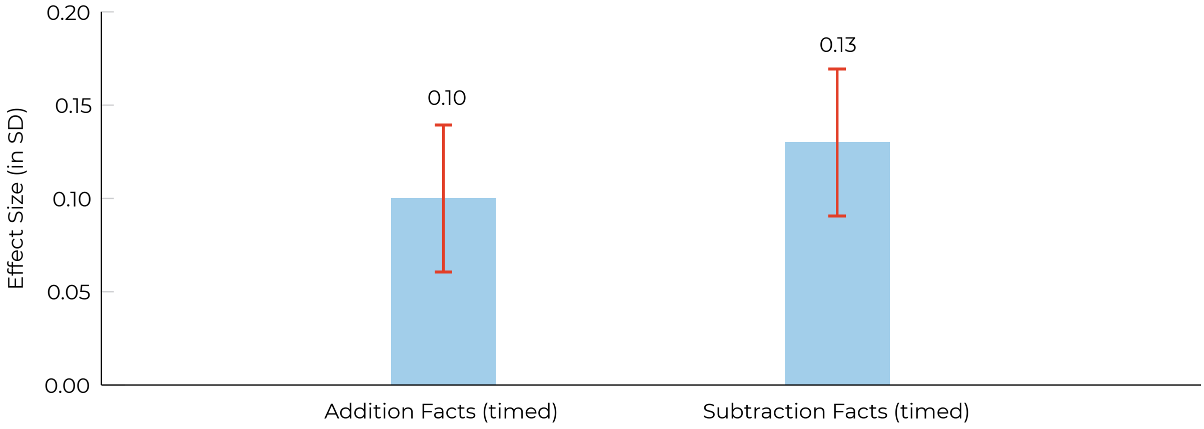

Figure 7: Task wise DiD Effect Sizes (in SD) – Timed Tasks.

The red lines denote the standard errors.

In contrast, skills such as Word Problems and Shape Recognition did not show statistically significant gains. This is likely linked to well-documented challenges: word problems require linguistic comprehension and the ability to translate text into mathematical structure[26], while shape recognition often receives limited classroom exposure, falls outside core number work and may show variability due to low student familiarity[27].

26 Verschaffel, L., Greer, B., & Van Dooren, W. (2020). Mathematical word problems: A critical analysis. Routledge.

27 Clements, D. H., & Sarama, J. (2009). Learning and teaching early math: The learning trajectories approach. Routledge. Clements and Sarama argue that early-grade mathematics instruction is dominated by number-focused activities, while geometry—particularly shape recognition—receives limited and often superficial attention. As a result, children frequently develop fragile, prototype-based understandings of shapes that depend strongly on instructional exposure. In programmes that prioritise number skills, this imbalance can lead to stagnation, or even apparent relative decline, in geometry-related outcomes compared with gains in other domains.

Overall, the results indicate relatively stronger improvements in numeracy skills related to basic number sense and fact fluency, with comparatively smaller changes observed in tasks requiring deeper reasoning or language processing. Gains across multiple skills suggest incremental progress across domains rather than concentrated effects in a limited set of areas. More limited movement in certain skills is consistent with the greater cognitive and instructional demands associated with these domains, which often develop over longer time periods.

4.2 Equivalent Years of Schooling (EYOS)

4.2.1 Methodology

The Equivalent Years of Schooling (EYOS) metric is used to express standardized numeracy learning gains in units of schooling-equivalent years. For Uttar Pradesh, EYOS is computed using data from the quasi-experimental evaluation, where the control group represents **business-as-usual (BAU)**[28] instruction.

Learning gains are estimated separately for intervention and control groups as the difference between endline and baseline mean numeracy scores. These gains are standardized using a pooled standard deviation to obtain effect sizes at the numeracy-competency level. EYOS is then computed using a **control-as-standard specification**, defined as the ratio of the intervention effect size to the control effect size. This ratio captures learning acceleration relative to the BAU learning trajectory over the same period.

EYOS estimates are aggregated across numeracy competencies using the below specifications:

- **Weighted EYOS**, where competencies are aggregated using different weights based on the average of intervention and control sample sizes across baseline and endline. This specification gives greater weight to competencies with larger and more reliable samples and is treated as the primary estimate.

Treatment of Extreme Values and Winsorization

Ratio-based EYOS estimates are sensitive to instability when control-group learning gains are small or heterogeneous across competencies, a concern that is particularly relevant in early-stage implementations

To ensure stability and interpretability of EYOS estimates for Uttar Pradesh, the following rules are applied:

- **Control effect size floor:** Competency-level observations with very small control effect sizes are excluded from ratio calculations to avoid inflated EYOS ratios driven by near-zero denominators.

28 Angrist, N., Bergman, P., Brewster, C., & Matsheng, M. (2020). Stemming learning loss during the pandemic: *A rapid randomized trial of a low-tech intervention in Botswana*. **Journal of Human Resources, 56**(S), S1–S45. https://doi.org/10.3368/jhr.58.S1.0620-12203R1 . **Business-As-Usual (BAU)** refers to the counterfactual learning trajectory that students would experience in the absence of the intervention, under prevailing instructional practices and system conditions. In impact evaluations, BAU is typically represented by outcomes observed in control or standard-implementation groups and serves as the benchmark against which incremental learning gains from an intervention are measured.

- **Upper-tail winsorization:** EYOS ratios are winsorized at an upper cap to limit the influence of extreme values arising from sparse competencies or noisy estimates. Values exceeding the cap are set equal to the cap rather than removed.
- **Consistent application:** Floors and caps are applied uniformly across weighted and unweighted specifications to maintain comparability.

Thresholds for floors and caps are selected based on inspection of the empirical distribution of effect sizes and are documented to support transparency and replicability. **To avoid denominator-driven inflation, control effect sizes below 0.20 SD were excluded and EYOS ratios were winsorized at 1.30×, consistent with early-stage implementation and higher variability.**

Uncertainty around EYOS estimates is quantified using standard errors derived via the delta method, enabling the construction of confidence intervals and hypothesis tests where feasible.

4.2.2 Findings

The EYOS analysis for TopParent in Uttar Pradesh indicates **positive learning gains in numeracy relative to BAU schooling**, though the magnitude of gains is more modest compared to some other contexts.

Specification	Intervention : Control EYOS Ratio	EYOS Above BAU (Δ years)	Learning Acceleration (%)	Interpretation
Weighted EYOS	1.37×	+0.3–0.4 years	+37%	Preferred estimate; reflects aggregation with sample-size-based difference weights

Under the **weighted specification**, the control-as-standard EYOS ratio is estimated at **1.37**, implying learning progress equivalent to a **37% increase** over the BAU trajectory during the evaluation period. In schooling-equivalent terms, this corresponds to approximately **0.3–0.4 additional years of learning** beyond BAU. This estimate is statistically distinguishable at the **90% confidence level**, but not at the 95% level and should therefore be interpreted with appropriate caution.

The EYOS results for Uttar Pradesh suggest early but meaningful numeracy learning gains attributable to TopParent. While estimates are less precise, the consistency across specifications indicates positive learning acceleration relative to BAU instruction.

4.3 Impact by Grade

This section examines performance across grades to understand how the intervention affected students at different stages of learning and to highlight grade-specific trends in numeracy skills.

The grade-wise analysis reveals a clear pattern: **Grade 2 shows gains across most foundational numeracy tasks, while Grade 3 demonstrates more uneven progress**, with smaller improvements in several areas and declines in others. This suggests that the intervention was comparatively **more effective in earlier grades**, particularly in strengthening number sense and arithmetic facts. One possible explanation is that foundational numeracy skills are still consolidating at this stage, allowing greater scope for improvement[29] and responsiveness to structured instructional support[30].

For Grade 2, positive gains were observed in 9 of the 11 skills assessed. The strongest improvements are seen in **Number Comparison** (0.16 SD) and **Subtraction Facts** (0.16 SD), both statistically significant and indicating strengthened number sense and basic fact retrieval. Subtraction Process (0.16 SD) also shows significant gains, along with improvements in Addition Facts (0.13 SD) and Addition Process (0.14 SD). The relatively higher growth in multi-step, procedural skills among Grade 2 students suggests that younger learners—who may have had less prior exposure to these concepts—are able to show greater gains, whereas Grade 3 students may have practised these skills more in school or may need to adjust previously learned approaches[31].

Counting in Bundles (0.08 SD)—a task requiring grouping and regrouping—shows smaller gains, indicating that place-value reasoning may not have strengthened at the same rate as fact fluency[32]. Grade 2 also shows mixed performance in Word Problems and Shape Recognition.

29 Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. American Economic Review, 97(2), 31–47. https://doi.org/10.1257/aer.97.2.31. Cunha and Heckman conceptualise skill formation as a dynamic, stage-specific process with "sensitive periods," in which early investments in foundational skills increase the productivity of later learning, while remediation at later stages remains possible but is typically more costly. This framework supports the view that early grades are more malleable and offer greater scope for improvement.

30 Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, 20(2), 82–88. https://doi.org/10.1016/j.lindif.2009.07.004; Doabler, C. T., Baker, S. K., Kosty, D. B., Smolkowski, K., Clarke, B., Miller, S. J., & Fien, H. (2015). Examining the association between explicit mathematics instruction and student mathematics achievement. The Elementary School Journal, 115(3), 303–333. https://doi.org/10.1086/679969. Jordan et al show that early *symbolic number sense* (counting, number knowledge, and basic arithmetic operations) measured at the start of Grade 1 uniquely predicts mathematics achievement in both Grades 1 and 3, even after accounting for age and broader cognitive abilities—supporting the idea that foundational numeracy is still consolidating in early grades and can be highly responsive to structured instructional inputs. Further, Doabler et al note that observational evidence from mathematics classrooms shows that the rate and quality of explicit instructional interactions (e.g., clear modeling, guided practice, feedback/checks for understanding) are positively associated with students' mathematics achievement—supporting the interpretation that early-grade learners can be particularly responsive to structured/explicit instruction while foundational skills are still developing.

31 Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539–1552. https://doi.org/10.1037/a0025510. Longitudinal evidence suggests that younger students often exhibit larger gains in procedural and multi-step mathematical skills because they are still actively developing foundational strategies, whereas older students tend to progress more slowly as they consolidate—or in some cases relearn—previously acquired approaches.

32 Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester Jr. (Ed.), Second handbook of research on mathematics teaching and learning (pp. 371–404). Information Age Publishing.Two-digit addition and subtraction can show improvement even when gains in Counting in Bundles are smaller because these operations rely more heavily on procedural steps and practice-based fluency. Students may learn to apply rules such as carrying and borrowing without fully internalising the underlying place-value structure of tens and ones. In contrast, Counting in Bundles directly assesses conceptual understanding of grouping and regrouping quantities, which typically develops more gradually.

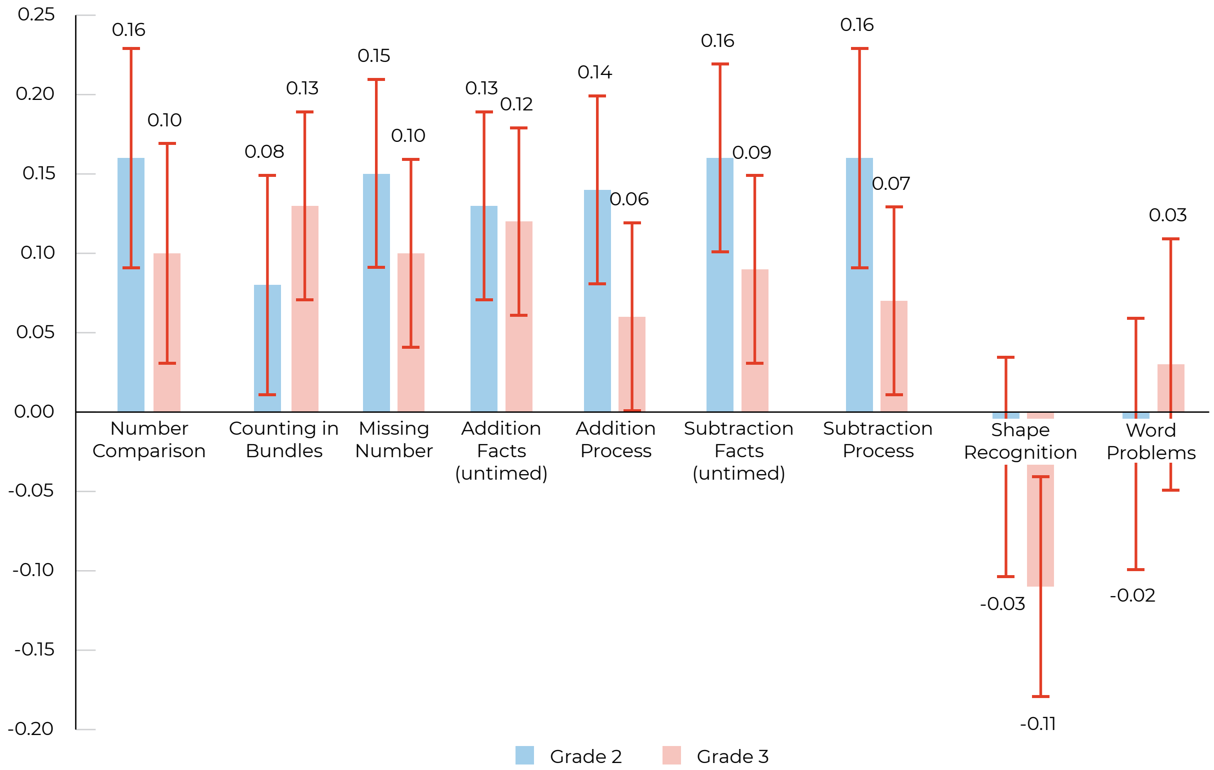

Figure 8: Grade-wise DiD Effect Sizes (in SD) – Untimed tasks.

The red lines denote the standard errors.

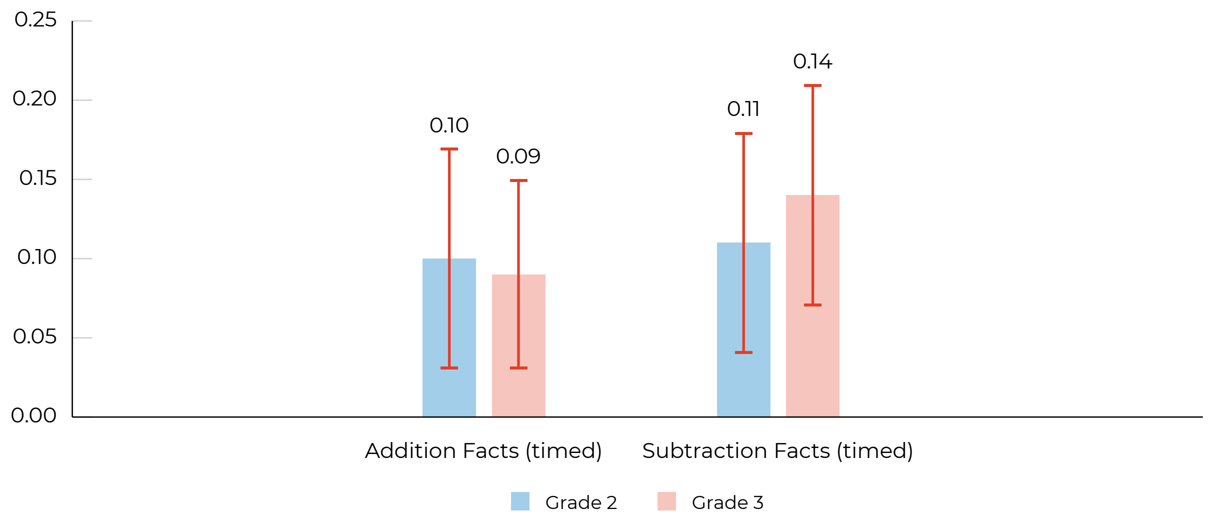

Figure 9: Grade-wise DiD Effect Sizes (in SD) – Timed tasks.

The red lines denote the standard errors.

For Grade 3, the pattern is more mixed. While there are **improvements in some procedural and conceptual tasks** - such as Addition Facts (untimed – 0.12 SD), Subtraction Process (0.07 SD) and Missing Number (0.10 SD) - **gains are generally smaller than for Grade 2**. Notably, Counting in Bundles (0.13 SD) shows moderate improvement for Grade 3 and is higher than for younger students. This suggests that higher grades may benefit more in this specific conceptual domain.

Timed tasks for Grade 3 reveal a slightly different trend: performance in Subtraction Facts (0.14 SD) is higher than Grade 2, suggesting some improvement in rapid retrieval or fluency under time pressure, though Addition Facts (0.09 SD) shows a smaller gain, similar to Grade 2.[33]

Overall, these patterns indicate that the programme had a stronger influence on procedural skills for Grade 2, with clearer gains in number sense, fact retrieval, and simple operations. Grade 3 shows more uneven progress, with gains focused in procedural tasks. These findings suggest that while the programme effectively reinforces early numeracy foundations, higher grades may require additional support to strengthen the reasoning and linguistic skills needed for conceptual tasks.

4.4 Impact by Gender

This section examines performance differences between boys and girls to identify potential gender gaps in learning outcomes and assess whether the programme benefits were equitably distributed[34].

The Top Parent intervention shows **stronger and more consistent gains for boys** in procedural skills. Girls demonstrate smaller but broadly positive improvements, with fewer sharp increases but also fewer declines. This suggests the programme supported both groups, but in different ways across the skill spectrum.

Boys recorded the largest gains in Addition and Subtraction Facts—0.23 SD and 0.15 SD (untimed) and 0.17 SD and 0.16 SD (timed) - reflecting **strong arithmetic accuracy and fluency**. They also showed gains in Subtraction Process (0.15 SD) and Addition Process (0.13 SD), reflecting progress in conceptual skills. However, performance was mixed across certain tasks - while Shape Recognition (0.01 SD) showed null effect, Counting in Bundles (0.17 SD) improved significantly. These areas require **grouping/regrouping, spatial reasoning, or multi-step comprehension - domains where boys displayed less consistent progress.**

33 Prior research indicates that higher-order numeracy tasks draw on additional cognitive and conceptual resources beyond basic computation. Arithmetic word problems require students to construct a situation model from text and translate it into a mathematical representation, placing demands on language comprehension, reasoning, and working memory (Hickendorff, 2021). Similarly, accurate execution of multi-digit column addition and subtraction depends on conceptual understanding of the base-10 system and place-value decomposition, which supports the reasoning required for multi-step procedures (Laski et al., 2016). Ref - Hickendorff, M. (2021). The demands of simple and complex arithmetic word problems on language and cognitive resources. Frontiers in Psychology, 12, 727761. https://doi.org/10.3389/fpsyg.2021.727761 ; Laski, E. V., Schiffman, J., Shen, C., & Vasilyeva, M. (2016). Kindergartners' base-10 knowledge predicts arithmetic accuracy concurrently and longitudinally. Learning and Individual Differences, 50, 234–239. https://doi.org/10.1016/j.lindif.2016.08.004

34 Boys and girls were not comparable in their numeracy skills at study onset, with only 2 out of 11 measures demonstrating gender equivalence ($p>0.05$): Addition Facts - untimed ($p = 0.09$) and Shape Recognition ($p = 0.75$).

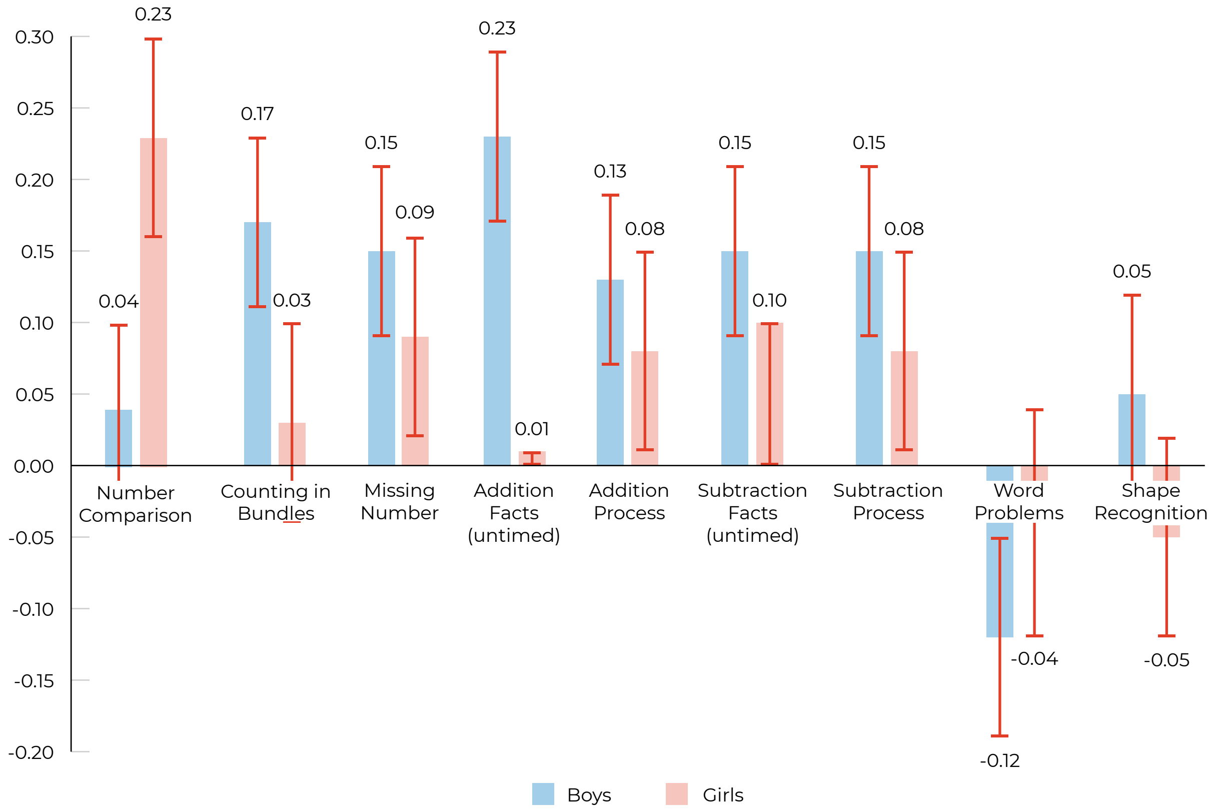

Figure 10: Gender-wise DiD Effect Sizes (in SD) – Untimed tasks.

The red lines denote the standard errors.

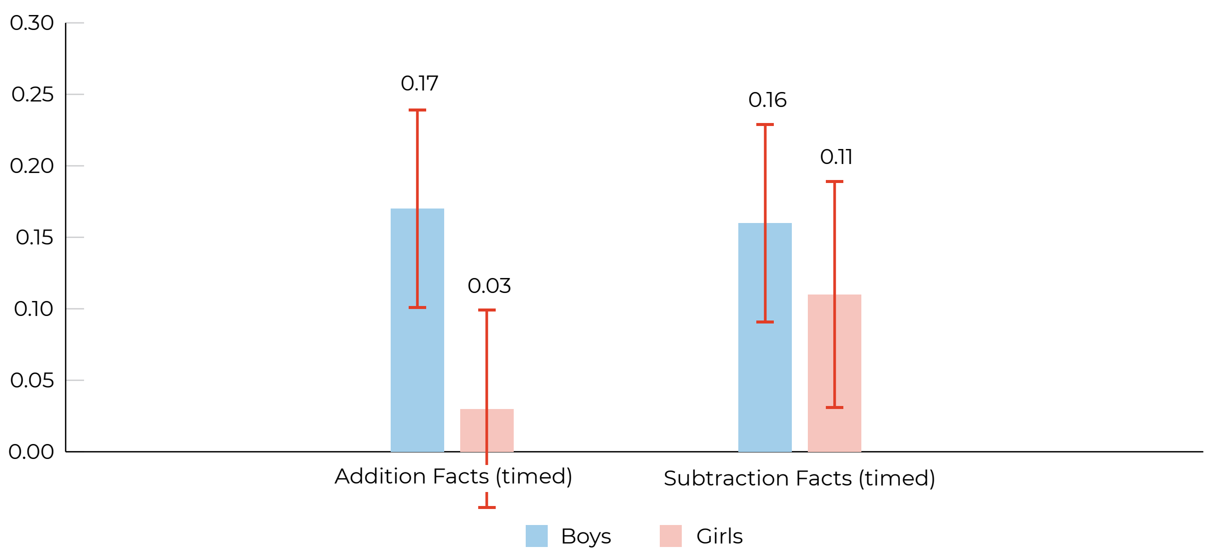

Figure 11: Gender-wise DiD Effect Sizes (in SD) – Timed tasks.

The red lines denote the standard errors.

Girls, in contrast, demonstrated **steady but smaller gains** across most tasks. Their strongest gains appeared in Number Comparison (0.23 SD) and Counting in Bundles (0.17 SD), both statistically significant, indicating **strengthened quantity sense and place-value understanding**. Gains in Missing Number (0.09 SD), Addition Process (0.08 SD) and

Subtraction Process (0.08 SD) show gradual improvement in procedural fluency. While girls recorded a small negative effect in Shape Recognition (-0.02 SD) and a decline in Word Problems (-0.12 SD), their overall pattern shows **fewer extreme drops compared to boys**. In timed tasks, girls also displayed progress in Subtraction Facts (0.11 SD), though gains in Addition Facts (0.03 SD) were minimal.

Overall, the results point to a gendered difference in how students responded to the intervention. Boys benefited in arithmetic facts and basic operations, but showed uneven performance in applied reasoning and spatial skills. Girls demonstrated improvements across most areas, particularly in number sense and other procedural tasks, though their gains were smaller in magnitude. This suggests that while the programme effectively strengthened foundational numeracy for both groups, boys may require targeted reinforcement in reasoning tasks and word problems, whereas girls may benefit from additional support in strengthening rapid arithmetic fluency.[35]

4.5 Score Distribution and Movement Analysis

This section presents the distribution of students across performance bands over assessment rounds, along with observed progressions and declines in student scores. Student performance at the sub-task level has been classified into five levels (L0–L4), defined separately for accuracy-based and fluency-based tasks, as outlined below.

Table 2: Categorisation for Accuracy & Fluency Tasks

Category	Accuracy Tasks (%)	Fluency Tasks (CPM - Correct Per Minute)
L0	0%	0
L1	0 to 25%	0 to 0.25 times Baseline Average (BL Avg.)
L2	25 to 50%	0.25 to 0.5 times BL Avg.
L3	50 to 75%	0.5 to BL Avg.
L4	75 to 100%	> BL Avg.

Note: All category lower limits are inclusive, except for L1; for example, 25–50% includes 25% but excludes 50%.

These categories enable tracking of the magnitude and direction of change in student performance across the evaluation timeline.

35 Successful performance in word problems depends not only on numerical skills but also on constructing meaning from text and translating linguistic information into a mathematical representation—making this task especially sensitive to learners' language comprehension and related reasoning resources (Hickendorff, 2021). Developmental evidence also indicates that girls, on average, demonstrate slightly earlier emerging language skills in early childhood (Eriksson et al., 2012). In contrast, large-scale assessment evidence using timed/basic computation indicators finds a small but systematic male advantage in arithmetic fluency measures (Räsänen et al., 2021). Taken together, this supports a targeted programme implication: boys may benefit from additional scaffolding for language-mediated reasoning tasks (including word problems), while girls may benefit from additional practice to strengthen rapid arithmetic fluency. Ref: Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovačević, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. British Journal of Developmental Psychology, 30(2), 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x; Hickendorff, M. (2021). The demands of simple and complex arithmetic word problems on language and cognitive resources. Frontiers in Psychology, 12, 727761. https://doi.org/10.3389/fpsyg.2021.727761; Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M.-J., & Korhonen, J. (2021). Effects of gender on basic numerical and arithmetic skills: Pilot data from third to ninth grade for a large-scale online dyscalculia screener. Frontiers in Education, 6, 683672. https://doi.org/10.3389/feduc.2021.683672.

The tables that follow present the percentage distribution of students across performance bands at Baseline and Endline for each assessed sub-task. Percentages reflect the proportion of students in each category at the respective assessment round, such that L0 + L1 + ... + L4 = 100% for Intervention and Comparison groups[36].

Table 3: Baseline Performance Category - Student Distribution

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Number Comparison (untimed)	17%	3%	7%	13%	60%	11%	1%	4%	12%	72%
Counting in Bundles (untimed)	44%	0%	15%	10%	31%	31%	0%	12%	13%	44%
Missing Number (untimed)	23%	8%	27%	22%	21%	11%	8%	25%	22%	33%
Addition facts (untimed)	21%	0%	7%	11%	61%	12%	0%	5%	11%	73%
Addition facts (timed)	17%	3%	6%	25%	48%	9%	3%	5%	19%	64%
Addition Process (untimed)	28%	11%	16%	29%	15%	18%	11%	15%	31%	25%
Subtraction facts (untimed)	24%	0%	13%	11%	52%	14%	0%	11%	10%	66%
Subtraction facts (timed)	24%	4%	8%	21%	43%	14%	3%	6%	17%	59%
Subtraction Process (untimed)	39%	10%	15%	25%	11%	26%	11%	14%	28%	21%
Word problem (untimed)	25%	25%	16%	23%	11%	18%	23%	19%	23%	18%
Shape Recognition (untimed)	85%	0%	14%	1%	0%	79%	0%	18%	3%	0%

From baseline to endline, the Intervention group shows a clear redistribution from L0–L2 into L3–L4 across most sub-tasks, with particularly strong gains in addition and subtraction facts (timed and untimed) where a majority of students move into L4. Comparison schools also improve, but remain more concentrated in mid-bands (L2–L3), especially on process and word-problem tasks.

Table 4: Endline Performance Category - Student Distribution

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Number Comparison (untimed)	15%	1%	6%	11%	68%	13%	1%	3%	8%	75%
Counting in Bundles (untimed)	28%	0%	14%	11%	47%	21%	0%	12%	12%	56%
Missing Number (untimed)	15%	7%	25%	22%	32%	12%	4%	22%	22%	39%
Addition facts (untimed)	15%	0%	5%	8%	71%	10%	0%	4%	8%	78%
Addition facts (timed)	13%	2%	5%	21%	59%	9%	1%	4%	17%	69%
Addition Process (untimed)	22%	11%	14%	32%	21%	17%	9%	15%	29%	30%
Subtraction facts (untimed)	20%	0%	11%	10%	59%	15%	0%	10%	9%	66%
Subtraction facts (timed)	21%	3%	7%	17%	53%	16%	3%	6%	15%	60%
Subtraction Process (untimed)	31%	12%	13%	28%	16%	26%	10%	13%	27%	25%
Word problem (untimed)	19%	21%	17%	24%	19%	14%	18%	15%	23%	30%
Shape Recognition (untimed)	81%	0%	17%	3%	0%	74%	0%	21%	5%	0%

36 Detailed counts of students in each category, disaggregated by Intervention and Comparison groups, are provided in Annexure 6.9.

In addition to examining the distribution of students across performance bands at Baseline and Endline, this study also tracked changes in individual student performance over time. As a longitudinal study involving the same students, this allows for an analysis of broad trends in movement across performance bands between the two rounds.

Across tasks, the score-transition patterns show that students in the **Intervention group exhibit stronger upward movement** between Baseline and Endline than those in the Comparison group. This is most evident among the lowest-performing students (L0): in Number Comparison (untimed), only 17% of Intervention L0 students remain at L0 while 67% jump directly to L4; in contrast, 20% of Comparison L0 students stay at L0 and only 52% reach L4. Similar differences appear in fluency tasks such as Addition Facts (untimed), where 60% of Intervention L0 students reach L4 compared with 49% in the Comparison group. Together, these shifts indicate that the intervention **accelerated movement out of the lowest band**.

Mid-level performers (L2 & L3) show an **even clearer divergence**. In Addition Facts (timed), over 80% of Intervention L2–L3 students progress to L3 or L4 (e.g., L2 → L4: 60%; L3 → L4: 69%), while the Comparison group displays a more modest rise (L2 → L4: 48%; L3 → L4: 53%). A similar pattern holds in Subtraction Facts (untimed), where 56–72% of Intervention L2–L3 students reach L4 compared to 49–64% in Comparison. These consistent differences indicate that Intervention students consolidate both fluency and accuracy more effectively over time, whereas Comparison students show flatter transitions with more stagnation at their baseline level.

Table 5: Score Movement Analysis - Numeracy[37]

Upward movewwment >10% Downward movement >10%

Number Comparison (untimed)					
Baseline Performance Category	Endline Performance Category				
	L0	L1	L2	L3	L4
Intervention					
L0	17%	2%	2%	11%	67%
L1	13%	0%	13%	13%	63%
L2	17%	2%	7%	17%	57%
L3	13%	3%	5%	15%	63%
L4	12%	1%	2%	7%	79%
Comparison					
L0	20%	2%	11%	16%	52%
L1	25%	0%	15%	23%	38%
L2	31%	1%	8%	16%	44%
L3	14%	2%	8%	17%	59%
L4	11%	1%	3%	7%	79%

Word Problems (Untimed)					
Baseline Performance Category	Endline Performance Category				
	L0	L1	L2	L3	L4
Intervention					
L0	23%	17%	17%	22%	20%
L1	17%	19%	14%	27%	23%
L2	8%	19%	17%	22%	34%
L3	10%	17%	15%	25%	33%
L4	13%	15%	10%	20%	41%
Comparison					
L0	36%	22%	14%	20%	7%
L1	17%	24%	19%	24%	16%
L2	14%	21%	18%	24%	23%
L3	11%	17%	18%	31%	23%
L4	11%	19%	14%	22%	34%

37 Detailed analysis has been shared in Annexure 6.9.1 and 6.9.2.

Performance shifts in conceptual tasks—such as Addition/Subtraction Process and Word Problems—are comparatively limited in both groups, reflecting their conceptual difficulty. Yet even here, the Intervention group maintains an advantage: in Word Problems, only 23% of Intervention L0 students remain at L0 compared to 36% in Comparison and a higher proportion reach L2–L3. High performers (L4) remain stable across tasks and groups.

The patterns indicate that mid-level students tend to benefit the most from supplemental support, while learners at the lowest performance levels show lesser improvement in the current structure. Taken together, these findings show that the intervention strengthened fluency and procedural skills, while deeper reasoning competencies demonstrated limited movement.

5. Discussion and Implications

The evaluation of the Top Parent numeracy intervention shows that the programme delivered consistent, modest gains across most foundational and mid-level numeracy skills, while also revealing distinct patterns across grades, gender and groups. Together, these findings reflect both the strengths of the programme model—particularly its focus on practice-based reinforcement of early number skills - and the challenges of shifting more complex reasoning competencies within the intervention window.

- **The EYOS estimate of approximately 1.4 years indicates that the intervention enabled students to achieve learning gains equivalent to almost one and a half years of Business-As-Usual instruction over the study period.** This magnitude reflects substantial learning acceleration rather than incremental improvement, suggesting that the programme meaningfully compresses the time required for students to reach expected competency levels. Such gains are particularly consequential in low-learning contexts, where delayed foundational skills compound over time. The EYOS result underscores the potential of well-designed, high-intensity EdTech models to not only raise learning levels but also close accumulated learning gaps within a relatively short timeframe, strengthening the case for scale-up where system capacity and implementation fidelity can be sustained.
- **Strengthening of foundational and fluency-based skills:** The programme generated gains in basic number sense, arithmetic fluency and other procedural tasks. Improvements in conceptual skills were comparatively less in magnitude, reflecting the greater cognitive demands of these skills possibly requiring longer intervention exposure[38]
- **Differential effects across learner groups** Grade 2 students demonstrated stronger and more consistent gains across domains, while Grade 3 learners showed flatter and more uneven progress. Gender-based patterns were also observed, with boys showing larger improvements in fluency and procedural skills, and girls demonstrating relatively greater gains in counting and number comparison.

38 Vessonen, T., Dahlberg, M., Hellstrand, H., Widlund, A., Söderberg, P., Korhonen, J., & Laine, A. (2025). The effectiveness of mathematical word problem-solving interventions among elementary schoolers: A systematic review and meta-analysis. *International Journal of Educational Research, 132*, 102642. Evidence from a systematic review and meta-analysis of elementary-level mathematical word-problem interventions indicates that dosage and frequency are significant moderators of effectiveness, suggesting that gains in conceptual problem-solving skills depend in part on the amount and regularity of instructional exposure (Vessonen et al., 2025).

- **Score movement aligned with effect-size patterns:** Students in the Intervention group, particularly those starting at the lowest performance levels, were more likely to transition into higher proficiency bands than their counterparts in the Comparison group. These shifts were most pronounced for fluency-based procedural tasks, while movement in conceptual skills remained more limited.

The evaluation highlights that TopParent meaningfully strengthens foundational numeracy skills and supports steady progress for low-performing students. While conceptual competencies show slower movement, the programme's design demonstrates potential for scalable, home-based numeracy support. Future iterations could improve gender-equitable outcomes, ensuring that TopParent continues to contribute effectively to early numeracy development in low-resource contexts.

6. Technical Annexures

6.1 Details of Sampling Design

- For the evaluation, using 95% confidence[39], 80% power4[40] and 0.1225 variance[41], the required sample size is 356 students each in the Intervention and Comparison groups.
- Minimal Detectable Effect Size (MDES)[42]: The study aims to be able to detect a 0.15 SD difference between the baseline and endline rounds. The formula used to calculate MDES is:

$$MDE = (t_K + t_\alpha) \cdot \sqrt{\frac{1}{P(1-P)} \cdot \frac{\sigma^2}{N}} \cdot \sqrt{1 + (m-1) \cdot ICC}$$

MDE = Minimal detectable effect
t_K and t_α = Critical values from Student's t for power K and significance level α
σ^2 = Variance; N = Sample size; P = Proportion in treatment; m = Cluster size;
ICC = Intraclass correlation

- Since sampling by school introduces intra-class correlation (ICC[43]), a design effect (DEFF) adjustment was applied. Assuming ICC = 0.1 and an average cluster size of 10 students, DEFF was calculated as:

$$DEFF = 1 + 0.1 \times (10 - 1) = 1.9$$

- This decreased the intervention and comparison sample size to 304 each[44].
- To account for absenteeism[45] (25% buffer), the sample increased to 405 per group. Factoring in 35% retention for a longitudinal study[46], the required sample became 1156 per group.
- Additionally, to estimate the Treatment-on-Treated (ToT) effect, it was assumed 30% of intervention students would meet the ideal usage threshold[47]. This required sampling 3854 intervention students.

6.2 Selection of Comparison Group

The comparison group for this quasi-experimental study would consist of schools that have not been given the treatment but are as similar as possible to the schools in the Intervention group in terms of baseline (pre-treatment) characteristics. The objective

39 The degree of certainty that the results reflect the true effect in the population rather than chance.
40 The probability of correctly detecting an effect if it truly exists.
41 The amount of natural variation in the outcome across individuals rather than uniform responses.
42 The smallest difference between groups that the study is designed to detect.
43 A measure of how similar students are within the same class or school, which affects the extent of independent information each student contributes. For this study, the intra-class correlation (ICC) was assumed to be 0.1, resulting in a Design Effect of 1.9.
44 Sample size decreases after applying Design Effect as there is a finite population for the intervention group.
45 The expected loss of sample due to students being absent during assessment, estimated at 25% for this study.
46 As the study was longitudinal and conducted over several months, some students present at baseline were unavailable at endline, hence an attrition buffer was included.
47 The estimated impact only on those who actually received the intervention as expected, estimated at 40% for this study.

of the comparison group is to capture what would have been the student learning outcomes in schools in the demonstration districts if the treatment was not given to these schools (i.e., the counterfactual). Hence, it is critical to identify the factors that are most likely to affect the outcome variable and then match the Intervention and Comparison observation units on these factors to create a valid counterfactual. While there are different techniques for creating a valid comparison group such as Propensity Score Matching, Regression Discontinuity Design and Mahalanobis Distance Matching, **Coarsened Exact Matching (CEM)** was used in this study for the following reasons:

1. A Monotonic Imbalance Bounding (MIB) matching method, CEM aims to balance the Intervention and Comparison groups ex ante
2. CEM also bounds through ex ante user choice both the average treatment effect estimation error and the degree of model dependence
3. CEM has been shown to produce good covariate balance between exposure groups and, thus, to reduce the impact of confounding in observational causal inference

6.2.1 Coarsened Exact Matching (CEM)[48]

The basic idea of CEM is to coarsen each variable by recoding it, so that substantially indistinguishable values are grouped and assigned the same numerical value. An exact matching algorithm is then applied to the coarsened data to determine the matches and to prune unmatched units.

In simple terms, the CEM process does not try to match each student in the Intervention group with a comparable student in the comparison group. Instead, it groups the characteristics of all the students in a particular school in the intervention group and then tries to find a school in the comparison group whose grouped characteristics exactly match the grouped characteristics of the original school. For example, let us assume that there is a school in the intervention group with 5 students in Grade 2, whose age (in years) and gender (M / F) is as follows:

- 6, M
- 6, M
- 6, M
- 7, F
- 7, F

The CEM method will coarsen this data into groups or bins and represent the school as one with less than 10 students in Grade 2, between 50% to 75% of whom are male and whose ages are either 6 or 7 years, with 50% to 75% of them being 6 years old. As a result, this school will be an exact match for the following school in the same district where the intervention is not being implemented and which has 9 students in Grade 2, whose age (in years) and gender (M / F) is as follows:

- 6, F
- 6, F

48 Iacus, S. M., King, G., & Porro, G. (2011). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1–24. https://doi.org/10.1093/pan/mpr013; Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493), 345–361. https://doi.org/10.1198/jasa.2011.tm09599.

- 6, F
- 6, F
- 6, M
- 6, M
- 7, M
- 7, M
- 7, M

Thus, the result of the CEM process in this case will be that each school in the sampling frame for the intervention group will be paired with a comparable school in the sampling frame for the comparison group and the sampling of the intervention and comparison groups will be done from these two sets of schools. To account for unforeseen challenges and situations on the ground during the data collection, a list of replacement schools will also be identified for the schools that make up the sampling frame for the comparison group.

6.2.2 Variables for the CEM Process:

Given below is the list of school-level variables that were used for the CEM process, to conduct the pairwise matching of schools, based on the data available on the UDISE website. These variables were chosen due to their potential relevance to student learning outcomes, informed by prior project experience.

Table 6: School-level variables

#	School-level variables
1	Total number of students in Grade 2
2	Total number of students in Grade 3
3	Pupil teacher ratio for the entire school
4	Location (urban vs. rural)
5	School (national) management type (e.g., dept. of education, local body, govt. aided, etc.)
6	Composite Index of Technological Infrastructure related variables (comprising of: count of functional laptops, desktops, tablets, scanners, printers, webcams, digi-boards, and access to the internet)[49]
7	School category (e.g., primary, primary with upper primary, etc.)

In scenarios where no exact match was present in the district for an intervention group school using all 7 variables mentioned above, some of these school-level variables were either coarsened further or dropped entirely in the subsequent rounds of the matching process, to ensure that a match was identified, while excluding the already matched intervention group schools. For example, in the absence of an exact match, the number of students in Grade 2 was coarsened into class intervals with a width of 20 students instead of 10 students for the next round of matching. However, the first 4 variables mentioned above were always retained in the CEM process.

49 All these variables are dichotomous in the UDISE dataset (meaning they have values of 1 for yes and 0 for no). The composite index was created by multiplying by taking the maximum value of the component variables, i.e., if any one of the variables had a value of 1, the composite variable would also have a value of 1.

6.2.3 Summary Table of Matching Exercise

Based on the approach detailed above, a summary of the results of the matching exercise is given below:

Table 7: Summary of Matching Exercise

Matching Round	No. of unique intervention group schools	No. of unique comparison group schools	Details of matching variables
1	147	123	The count of students in Grades 2 & 3 and the pupil-teacher ratio, were coarsened into class intervals of width 10. All 147 Intervention groups schools matched in this round could have been covered by 78 Comparison group schools, but the number was increased to 123 using a many-to-many matching approach, to sample the minimum number of Comparison group schools required to maintain a cluster size of at least 10.
2	40	10	The school category variable was dropped. All 40 Intervention groups schools matched in this round could have been covered by 5 additional unique Comparison group schools, along with 17 Comparison group schools from the previous round.
3	26	15	The aggregated tech infra variable was also dropped. All 26 Intervention groups schools matched in this round could have been covered by 7 additional unique Comparison group schools, along with 11 Comparison group schools from the previous rounds.
4	23	10	The national management variable was also dropped. All 23 Intervention groups schools matched in this round could have been covered by 5 additional unique Comparison group schools, along with 13 Comparison group schools from the previous rounds.
5	60	22	The location variable was also dropped. All 60 Intervention groups schools matched in this round could have been covered by 15 additional unique Comparison group schools, along with 15 Comparison group schools from the previous rounds.
6	35	12	The pupil-teacher ratio variable was also dropped. All 35 Intervention groups schools matched in this round could have been covered by 10 additional unique Comparison group schools, along with 15 Comparison group schools from the previous rounds.
7	2	0	All the dropped variables were re-introduced and the width of the class intervals for the coarsened count of students in Grades 2 and 3 was increased to 125. Additionally, the width of the class interval for the pupil-teacher ratio was also increased to 25.
8	3	0	The school category variable was dropped again.

Matching Round	No. of unique intervention group schools	No. of unique comparison group schools	Details of matching variables
9	1	2	The aggregated tech infra and national management variables were also dropped again.
10	2	2	The location variable was also dropped again.
11	2	3	The pupil-teacher ratio variable was also dropped again.
Total	**341**	**199**	

The final matching process paired all 139 intervention group schools with 64 comparison group schools. One comparison group school was matched with up to 40 intervention schools and one intervention school with up to 35 comparison schools. While technically 32 comparison schools would have covered all intervention schools, 64 comparison schools were selected to maintain an average cluster size of 15 students per school, leading to a comparison group sample of 960 students.

6.3 Final Sample Achieved

The table below presents a detailed summary of the number of students assessed as part of the study, along with the number excluded from the analysis due to the following reasons:

- Attempting only one subject in a given round due to lack of consent, operational challenges, or other constraints
- Absent or unsynced audio response recordings / assets
- Presence in the Baseline but absence in the Endline

Table 8: Final Sample Achieved in Intervention and Comparison groups

Description	Intervention				Comparison			
	Baseline		Endline		Baseline		Endline	
	Literacy	Numeracy	Literacy	Numeracy	Literacy	Numeracy	Literacy	Numeracy
Total number of assessments administered	3414	3381	2766	2761	1667	1657	1254	1258
Students excluded due to operational constraints[50]	90	30	57	25	20	15	10	19
Students excluded due to technical issues[51]	9		101		109		103	
Students excluded as per study design criteria[52]	1424		744		412		10	
Final Student Sample	1891				1126			

50 Students who attempted only one of the two subjects within a given assessment round.
51 The student's assessment having absent or unsynced recordings/assets.
52 Students not present in both assessment rounds i.e. Baseline and Endline.

6.4 Modified EGMA Tool

The below table summarises the tool utilised in this evaluation.

Table 9: Details of modified Numeracy Tool and Skills Covered

Concept / Skill Assessed	Total Items/ Questions	Nature of Task	Metric	Objective	Mode of Evaluation
Number discrimination (Identify and read out the greater number among the given pairs of numbers)	10	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can compare and order numbers up to 1,000 accurately.	Automated
Counting in bundles (Give the total count of straws when given a few bundles of 10 straws each, along with some single straws. Also choose the correct count of bundles of 10 straws and single straws that add up to a given number.)	4	Untimed	Number of questions answered correctly	The objective of this task is to check if the child understands the concept of place value, i.e., if they understand how many tens and ones make a 2-digit number.	Automated
Number patterns (Identify a pattern in numbers and call out the missing number in each pattern)	8	Untimed	Number of missing numbers identified correctly	The objective of this task is to check if the child can recognize and complete patterns involving skip counting in 1s, 10s, 5s and 2s, for numbers up to 1,000.	Automated
Single-digit addition facts (Solve single-digit addition problems)	3	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can add single-digit numbers fluently.	Automated
	20	Timed: 1 minute			Automated
Multi-digit addition in vertical format (Solve multi-digit addition problems)	5	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can add multi-digit numbers accurately.	Automated
Subtraction facts within 18 (Solve subtraction problems involving numbers within 18 and with single-digit answers)	3	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can subtract numbers up to 18 and with single-digit answers, fluently.	Automated
	20	Timed: 1 minute			Automated

Concept / Skill Assessed	Total Items/ Questions	Nature of Task	Metric	Objective	Mode of Evaluation
Multi-digit subtraction in vertical format (Solve multi-digit subtraction problems)	5	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can subtract multi-digit numbers accurately.	Automated
Operations involving 0 (Solve addition and subtraction problems involving 0)	4	Untimed	Number of questions answered correctly	The objective of this task is to check if the child understands the concept of zero accurately. 1 question each in addition facts (untimed), addition process, subtraction facts (untimed) and subtraction process[53].	Automated
Word problems on single-digit number operations (Give an answer to narrated word problems)	6	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can apply basic addition and subtraction in real-world scenarios.	Automated
Shape recognition **(Identify a specific shape among a collection of shapes)**	3	Untimed	Number of shapes recognized correctly	The objective of this task is to check if the child can recognize basic shapes like circles, rectangles and triangles. One question each on circles, rectangles and triangles. Each question involves showing the student multiple shapes (8, 10 and 11 respectively), of which only a few (2, 4 and 4 respectively) are correct examples of the shape the student has been asked to identify.	Automated

53 These items are already included in the count of the respective sub-tasks.

6.5 Demographic Analysis of Students Who Dropped Out after Baseline

Table 10: Demographic Analysis of Baseline Dropouts

Indicator		Intervention		Comparison	
		Students absent after Baseline	Overall Student Pool	Students absent after Baseline	Overall Student Pool
No. of Students		1890	171	532	667
Student Age (avg.)		9.3 (9.2 - 9.4)	9.3 (9.2 - 9.4)	9.3 (9.0 - 9.6)	9.2 (9.1 - 9.3)
Grade Split	Class 2	53% (43% - 61%)	53% (49% - 57)	50% (39% - 60%)	46% (35% - 56%)
	Class 3	47% (40% - 54%)	47% (44% - 51%)	50% (37% - 64%)	54% (40% - 68%)
No. of Sisters (avg.)		1.6 (1.5 - 1.6)	1.6 (1.5 - 1.6)	1.8 (1.6 - 1.9)	1.7 (1.6 - 1.8)
No. of Brothers(avg.)		1.6 (1.5 - 1.7)	1.6 (1.6 - 1.7)	1.8 (1.7 - 1.9)	1.7 (1.7 - 1.8)
Socio-economic Context**	Low	3% (-4% - 9%)	2% (0% - 5%)	7% (-4% - 18%)	8% (-4% - 19%)
	High	97% (96% - 99%)	98% (97% - 98%)	93% (90% - 95%)	92% (90% - 95%)

* 95% Confidence Interval mentioned in brackets
** Binary flag created on the basis of appliances/transport available at home (i.e. feature phone v. smart phone;cycle v. motor cycle)

Across both Intervention and Comparison groups, students who were absent after baseline were demographically similar to the overall student pool on key characteristics, with point estimates generally close and 95% confidence intervals largely overlapping.

In the Intervention group, the average age of students absent after baseline was identical to the overall pool (9.3 years; CI: 9.2–9.4 for both). In the Comparison group, absent students were also of a similar age (9.3 years; CI: 9.0–9.6) compared to the overall pool (9.2 years; CI: 9.1–9.3), with overlapping intervals suggesting only a small difference at most. Grade composition was likewise comparable: in the Intervention group, absentees were split evenly across Class 2 and 3 in the same proportions as the overall pool (Class 2: 53% vs. 53%; Class 3: 47% vs. 47%, with overlapping confidence intervals). In the Comparison group, absentees showed a slightly higher share in Class 2 (50%; CI: 39%–60%) relative to the overall pool (46%; CI: 35%–56%), though the wide and overlapping intervals indicate this difference is modest and uncertain.

Household composition, measured through the average number of sisters and brothers, was closely aligned between students absent after baseline and the overall pool. In the Intervention group, averages were effectively the same for both sisters (1.6; CI: 1.5–1.6 for both) and brothers (1.6; CI: 1.5–1.7 among absentees vs. 1.6; CI: 1.6–1.7 overall). In the Comparison group, absent students reported slightly higher sibling counts on average—1.8 sisters (CI: 1.6–1.9) and 1.8 brothers (CI: 1.7–1.9)—compared to the overall pool (1.7 sisters, CI: 1.6–1.8; 1.7 brothers, CI: 1.7–1.8). However, the confidence intervals overlap, suggesting only minor differences.

Socio-economic status patterns among students absent after baseline were also broadly consistent with the overall pool, especially given the wide uncertainty ranges for the "low" category. In the Intervention group, the proportion of absentees classified as low socio-economic status was 3% (CI: –4% to 9%), close to the overall pool estimate of 2% (CI: 0%–5%) and the corresponding high socio-economic shares were similarly aligned (97%, CI: 96%–99% among absentees vs. 98%, CI: 97%–98% overall). In the Comparison group, low socio-economic estimates were also close (7%, CI: –4% to 18% among absentees vs. 8%, CI: –4% to 19% overall), with high socio-economic shares nearly identical (93%, CI: 90%–95% vs. 92%, CI: 90%–95%). Overall, these patterns suggest that baseline absence was not strongly concentrated in any particular socio-economic segment.

Taken together, the demographic profile of students absent after baseline appears broadly similar to that of the overall study sample across age, grade composition, sibling counts and socio-economic context, with any observed differences small relative to the uncertainty in the estimates[54].

6.6 Overall Performance

Table 11: Detailed Results

Top Parent											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	74%	79%	69%	78%	5%	0%	36%	0.13	0.00
Counting in Bundles	Untimed	%	51%	59%	36%	48%	15%	11%	40%	0.11	0.01
Missing Number	Untimed	%	49%	56%	41%	51%	9%	5%	33%	0.12	0.01
Addition Facts	Untimed	%	72%	78%	63%	74%	9%	4%	37%	0.13	0.00
Addition Facts	Timed	CPM	8.48	9.90	6.87	8.83	1.61	1.07	5.30	0.10	0.00
Addition Process	Untimed	%	44%	51%	37%	47%	7%	4%	33%	0.10	0.02
Subtraction Facts	Untimed	%	62%	69%	56%	68%	6%	1%	39%	0.13	0.02
Subtraction Facts	Timed	CPM	6.7	7.9	5.4	7.3	1.25	0.60	5.15	0.13	0.00
Subtraction Process	Untimed	%	37%	44%	30%	41%	6%	2%	33%	0.12	0.01
Word Problems	Untimed	%	40%	50%	32%	39%	8%	10%	32%	0.00	0.14
Shape Recognition	Untimed	%	7%	10%	5%	8%	2%	2%	16%	-0.08	0.89

54 It is important to note that demographic information was not mandatory during data collection and therefore, while representative, it may be incomplete.

6.7 Grade-wise Results

Table 12: Detailed Results - Grade 2

Top Parent - Grade 2											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	70%	77%	63%	77%	7%	1%	38%	0.16	0.02
Counting in Bundles	Untimed	%	45%	55%	30%	44%	15%	12%	40%	0.08	0.19
Missing Number	Untimed	%	45%	54%	36%	49%	9%	4%	33%	0.15	0.02
Addition Facts	Untimed	%	67%	76%	57%	71%	10%	5%	39%	0.13	0.04
Addition Facts	Timed	CPM	7.77	9.59	5.95	8.32	1.82	1.28	5.30	0.10	0.08
Addition Process	Untimed	%	39%	47%	30%	43%	9%	4%	35%	0.14	0.03
Subtraction Facts	Untimed	%	56%	67%	49%	66%	8%	1%	40%	0.16	0.01
Subtraction Facts	Timed	CPM	5.97	7.69	4.69	6.99	1.28	0.70	5.15	0.11	0.05
Subtraction Process	Untimed	%	32%	40%	24%	37%	8%	3%	31%	0.16	0.01
Word Problems	Untimed	%	36%	47%	29%	39%	7%	8%	31%	-0.03	0.84
Shape Recognition	Untimed	%	7%	10%	5%	8%	2%	2%	16%	-0.02	0.94

Table 13: Detailed Results - Grade 3

Top Parent - Grade 3											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	78%	80%	75%	80%	4%	0%	34%	0.10	0.12
Counting in Bundles	Untimed	%	58%	62%	43%	52%	15%	10%	40%	0.13	0.03
Missing Number	Untimed	%	55%	58%	46%	52%	8%	5%	32%	0.10	0.11
Addition Facts	Untimed	%	77%	79%	70%	77%	7%	3%	35%	0.12	0.04
Addition Facts	Timed	CPM	9.29	10.15	7.91	9.24	1.38	0.91	5.21	0.09	0.14

Top Parent - Grade 3											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value
			I	C	I	C	I	C			
Addition Process	Untimed	%	49%	54%	44%	50%	5%	3%	34%	0.06	0.31
Subtraction Facts	Untimed	%	68%	70%	64%	69%	4%	1%	37%	0.09	0.17
Subtraction Facts	Timed	CPM	7.44	8.09	6.22	7.56	1.21	0.52	5.07	0.14	0.04
Subtraction Process	Untimed	%	42%	47%	38%	45%	4%	2%	34%	0.07	0.27
Word Problems	Untimed	%	45%	52%	36%	40%	8%	12%	32%	-0.11	0.09
Shape Recognition	Untimed	%	8%	10%	5%	8%	3%	2%	16%	0.03	0.81

6.8 Gender-wise Results

Table 14: Detailed Results - Boys

Top Parent - Boys											
Task	Task Type	Unit	Endline - Boys		Baseline - Boys		Delta - Boys		Pooled SD	Effect Size	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	74%	82%	71%	80%	3%	2%	36%	0.04	0.46
Counting in Bundles	Untimed	%	54%	61%	39%	53%	15%	8%	40%	0.17	0.00
Missing Number	Untimed	%	51%	57%	41%	53%	10%	5%	32%	0.15	0.01
Addition Facts	Untimed	%	73%	77%	64%	76%	9%	1%	37%	0.23	0.00
Addition Facts	Timed	CPM	8.89	10.20	7.09	9.27	1.81	0.93	5.08	0.17	0.01
Addition Process	Untimed	%	45%	51%	38%	48%	7%	3%	32%	0.13	0.04
Subtraction Facts	Untimed	%	64%	70%	58%	70%	6%	0%	38%	0.15	0.01
Subtraction Facts	Timed	CPM	6.94	8.16	5.64	7.62	1.30	0.54	4.81	0.16	0.02
Subtraction Process	Untimed	%	38%	45%	32%	44%	7%	2%	33%	0.15	0.01
Word Problems	Untimed	%	41%	52%	34%	41%	7%	10%	30%	-0.12	0.08
Shape Recognition	Untimed	%	7%	10%	5%	8%	2%	2%	16%	0.05	0.79

Table 15: Detailed Results - Girls

Top Parent - Girls											
Task	Task Type	Unit	Endline - Girls		Baseline - Girls		Delta - Girls		Pooled SD	Effect Size	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	74%	75%	66%	76%	8%	-1%	38%	0.23	0.00
Counting in Bundles	Untimed	%	48%	58%	32%	43%	15%	14%	38%	0.03	0.65
Missing Number	Untimed	%	47%	54%	40%	49%	8%	5%	32%	0.09	0.14
Addition Facts	Untimed	%	70%	79%	63%	72%	7%	7%	39%	0.01	0.82
Addition Facts	Timed	CPM	8.00	9.53	6.61	8.28	1.38	1.25	4.68	0.03	0.62
Addition Process	Untimed	%	42%	50%	35%	45%	7%	5%	33%	0.08	0.20
Subtraction Facts	Untimed	%	59%	67%	53%	64%	6%	3%	39%	0.10	0.14
Subtraction Facts	Timed	CPM	6.32	7.62	5.12	6.93	1.19	0.68	4.57	0.11	0.11
Subtraction Process	Untimed	%	35%	42%	29%	39%	6%	3%	32%	0.08	0.19
Word Problems	Untimed	%	39%	47%	30%	37%	9%	10%	29%	-0.04	0.76
Shape Recognition	Untimed	%	8%	10%	6%	7%	2%	3%	17%	-0.05	0.93

6.9 Score Movement Analysis

6.9.1 Baseline Performance Distribution - Count of Students

Table 16: Baseline Performance Category - Count of Students

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Number Comparison (untimed)	314	53	141	239	1142	128	9	42	135	810
Counting in Bundles (untimed)	829	0	283	196	581	350	0	131	151	490
Missing Number (untimed)	425	150	504	412	397	127	95	280	246	371
Addition facts (untimed)	401	0	127	202	1158	136	0	51	120	810
Addition facts (timed)	329	66	114	479	903	102	30	55	215	724

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Addition Process (untimed)	536	213	306	556	277	197	120	171	348	280
Subtraction facts (untimed)	448	0	243	207	990	151	0	117	113	733
Subtraction facts (timed)	460	77	146	397	811	157	38	72	194	665
Subtraction Process (untimed)	730	193	278	472	215	287	125	157	312	233
Word problem (untimed)	467	475	310	433	203	200	254	212	252	196
Shape Recognition (untimed)	1607	0	263	17	1	882	0	196	33	2

6.9.2 Endline Performance Distribution - Count of Students

Table 17: Endline Performance Category - Count of Students

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Number Comparison (untimed)	280	19	106	198	1280	146	13	29	95	838
Counting in Bundles (untimed)	534	1	265	201	879	232	0	132	129	625
Missing Number (untimed)	289	126	461	407	594	136	50	243	249	440
Addition facts (untimed)	286	0	101	151	1338	109	0	47	94	868
Addition facts (timed)	246	45	86	387	1110	99	13	42	188	776
Addition Process (untimed)	405	200	264	597	398	193	95	169	322	337
Subtraction facts (untimed)	367	0	200	187	1102	164	0	110	104	734
Subtraction facts (timed)	381	50	139	307	978	176	29	70	166	671
Subtraction Process (untimed)	564	219	242	518	304	287	106	139	304	275
Word problem (untimed)	359	384	310	449	341	157	196	164	259	334
Shape Recognition (untimed)	1480	0	305	48	4	823	0	232	53	1

6.9.3 Student Score Movement from Baseline to Endline - Intervention

Table 18: Score Movements - Intervention

>10% Downward Movement   >10% Upward Movement

Task	Baseline Performance Category	Endline Performance Category				
		L0	L1	L2	L3	L4
Number Comparison (untimed)	L0	20%	2%	11%	16%	52%
	L1	25%	0%	15%	23%	38%
	L2	31%	1%	8%	16%	44%
	L3	14%	2%	8%	17%	59%
	L4	11%	1%	3%	7%	79%
Counting in Bundles (untimed)	L0	43%	0%	14%	9%	33%
	L1	-	-	-	-	-
	L2	23%	0%	19%	14%	44%
	L3	13%	0%	14%	12%	61%
	L4	15%	0%	12%	10%	63%
Missing Number (untimed)	L0	32%	12%	24%	15%	17%
	L1	18%	12%	30%	17%	23%
	L2	13%	6%	30%	24%	28%
	L3	10%	3%	24%	28%	35%
	L4	6%	4%	17%	21%	52%
Addition facts (untimed)	L0	33%	0%	8%	10%	49%
	L1	-	-	-	-	-
	L2	21%	0%	7%	11%	60%
	L3	18%	0%	8%	11%	63%
	L4	8%	0%	4%	7%	82%
Addition facts (timed)	L0	31%	4%	8%	21%	36%
	L1	23%	2%	5%	15%	56%
	L2	13%	5%	10%	24%	48%
	L3	9%	3%	5%	30%	53%
	L4	8%	1%	2%	15%	72%

Task	Baseline Performance Category	Endline Performance Category				
		L0	L1	L2	L3	L4
Addition Process (untimed)	L0	40%	15%	12%	22%	11%
	L1	23%	11%	17%	34%	15%
	L2	14%	10%	17%	38%	21%
	L3	13%	10%	15%	38%	24%
	L4	10%	6%	11%	32%	41%
Subtraction facts (untimed)	L0	40%	0%	17%	10%	33%
	L1	-	-	-	-	-
	L2	26%	0%	13%	12%	49%
	L3	17%	0%	7%	12%	64%
	L4	10%	0%	8%	9%	73%
Subtraction facts (timed)	L0	38%	4%	10%	15%	32%
	L1	26%	8%	6%	21%	39%
	L2	28%	1%	11%	21%	39%
	L3	12%	2%	8%	20%	58%
	L4	13%	2%	5%	15%	66%
Subtraction Process (untimed)	L0	49%	11%	11%	19%	10%
	L1	26%	17%	18%	27%	12%
	L2	18%	13%	19%	34%	16%
	L3	18%	10%	13%	38%	21%
	L4	15%	11%	10%	30%	33%
Word Problems (untimed)	L0	36%	22%	14%	20%	7%
	L1	17%	24%	19%	24%	16%
	L2	14%	21%	18%	24%	23%
	L3	11%	17%	18%	31%	23%
	L4	11%	19%	14%	22%	34%
Shape Recognition (untimed)	L0	82%	0%	15%	2%	0%
	L1	-	-	-	-	-
	L2	73%	0%	23%	5%	1%
	L3	73%	0%	20%	7%	0%
	L4	100%	0%	0%	0%	0%

6.9.4 Student Score Movement from Baseline to Endline - Comparison

Table 19: Score Movements – Comparison

>10% Downward Movement   >10% Upward Movement

Task	Baseline Performance Category	Endline Performance Category				
		L0	L1	L2	L3	L4
Number Comparison (untimed)	L0	17%	2%	2%	11%	67%
	L1	13%	0%	13%	13%	63%
	L2	17%	2%	7%	17%	57%
	L3	13%	3%	5%	15%	63%
	L4	12%	1%	2%	7%	79%
Counting in Bundles (untimed)	L0	27%	0%	16%	12%	45%
	L1	-	-	-	-	-
	L2	22%	0%	13%	12%	53%
	L3	17%	0%	12%	15%	56%
	L4	17%	0%	9%	10%	65%
Missing Number (untimed)	L0	29%	10%	21%	19%	21%
	L1	13%	8%	29%	22%	27%
	L2	12%	4%	25%	26%	33%
	L3	8%	3%	24%	24%	40%
	L4	10%	3%	15%	19%	53%
Addition facts (untimed)	L0	25%	0%	7%	8%	60%
	L1	-	-	-	-	-
	L2	10%	0%	6%	12%	73%
	L3	12%	0%	5%	11%	73%
	L4	7%	0%	4%	8%	81%
Addition facts (timed)	L0	18%	2%	8%	21%	51%
	L1	33%	0%	17%	23%	27%
	L2	13%	2%	5%	20%	60%
	L3	7%	1%	3%	20%	69%
	L4	7%	1%	3%	15%	75%
Addition Process (untimed)	L0	32%	15%	19%	22%	12%
	L1	18%	11%	16%	31%	24%
	L2	17%	7%	18%	30%	28%
	L3	14%	8%	13%	33%	32%
	L4	10%	4%	13%	27%	45%

Task	Baseline Performance Category	Endline Performance Category				
		L0	L1	L2	L3	L4
Subtraction facts (untimed)	L0	26%	0%	17%	11%	46%
	L1	-	-	-	-	-
	L2	25%	0%	11%	8%	56%
	L3	20%	0%	11%	5%	64%
	L4	10%	0%	8%	10%	72%
Subtraction facts (timed)	L0	27%	6%	9%	19%	38%
	L1	32%	5%	0%	11%	51%
	L2	16%	1%	13%	11%	59%
	L3	15%	2%	7%	17%	59%
	L4	12%	2%	5%	14%	67%
Subtraction Process (untimed)	L0	39%	11%	16%	20%	15%
	L1	25%	13%	8%	35%	19%
	L2	21%	12%	16%	32%	19%
	L3	23%	7%	10%	34%	27%
	L4	16%	8%	11%	23%	42%
Word Problems (untimed)	L0	23%	17%	17%	22%	20%
	L1	17%	19%	14%	27%	23%
	L2	8%	19%	17%	22%	34%
	L3	10%	17%	15%	25%	33%
	L4	13%	15%	10%	20%	41%
Shape Recognition (untimed)	L0	76%	0%	19%	5%	0%
	L1	-	-	-	-	-
	L2	74%	0%	20%	5%	1%
	L3	69%	0%	28%	3%	0%
	L4	100%	0%	0%	0%	0%

6.10 Analysis of Literacy Investigation

6.10.1 Overview

The literacy assessment was administered alongside the numeracy assessment and covered an identical sample of students. The final analytical sample includes only those students who participated in both literacy and numeracy assessments across the Baseline and Endline rounds. Literacy outcomes were measured using a contextualised version of the Early Grade Reading Assessment (EGRA), which was administered uniformly to students in Grades 2 and 3 and remained consistent across both rounds. To ensure alignment with the programme context, the assessment tool was adapted and contextualised and translated into Hindi.

This literacy investigation is indicative and should be interpreted with appropriate caution. All the literacy sub-tasks relied on audio-based assessments. Variations in audio clarity and inconsistencies in recording conditions have influenced the measurement precision.

6.10.2 Indicative Investigation

The Top Parent literacy assessments show small, mostly non-significant changes across outcomes, with programme effects concentrated in early letter knowledge and less on broader reading fluency or comprehension.

The only statistically significant improvement is in Letter Recognition (untimed - 0.19 SD). Letter Recognition (timed) also shows a positive pattern (0.10 SD), suggesting some emerging gains in speeded recognition

Listening Comprehension (0.07 SD) and Oral Vocabulary (0.05 SD) demonstrated smaller, statistically insignificant gains. Word-level reading outcomes were also small and statistically insignificant - Familiar Word Reading (untimed - 0.08 SD; timed - 0.03 SD) and Nonword Reading (0.04 SD), suggesting minimal transfer from letter gains into decoding and automatic word reading within the evaluation period.

Table 20: Top Parent Literacy Results

Top Parent - Literacy											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value[55]
			I	C	I	C	I	C			
Listening Comprehension	Untimed	%	43%	44%	37%	41%	6%	3%	37%	0.07	0.16
Oral Vocabulary	Untimed	%	61%	60%	61%	62%	0%	-2%	27%	0.05	0.27
Letter Recognition	Untimed	%	67%	71%	64%	75%	3%	-4%	34%	0.19	0.00
Letter Recognition	Timed	CPM	26.1	30.5	23.5	29.6	2.6	0.9	17.4	0.10	0.07
Familiar Word Reading	Untimed	%	58%	67%	55%	66%	3%	1%	38%	0.08	0.12
Familiar Word Reading	Timed	CPM	16.3	20.8	14.0	18.9	2.3	1.9	13.9	0.03	0.61
Nonword Reading	Timed	CPM	15.1	19.2	12.9	17.5	2.3	1.7	13.8	0.04	0.45
Oral Reading Fluency	Timed	CPM	33.6	46.5	30.7	42.6	3.0	3.8	42.4	-0.02	0.69
Reading Comprehension	Untimed	%	37%	46%	30%	42%	7%	4%	36%	0.05	0.36

Overall, the pattern suggests the programme may be most effective at strengthening basic letter recognition, with limited measurable impact on conceptual reading outcomes such as fluency and comprehension in this sample.

55 A p-value of less than 0.05 is considered statistically significant.

LiftEd Transforming Foundational Literacy and Numeracy

EDTECH ACCELERATOR

To learn more about the

LiftEd EdTech Accelerator,

visit our website:

www.edtechaccelerator.org

Contact us at:
edtechaccelerator@centralsuarefoundation.org