JANUARY 2026

Evidence

The Limits of a "Great Buy": **What Rigorous Evidence Reveals about Teaching at the Right Level**

Report prepared by:
Cally Ardington (AFLEARN, University of Cape Town)
Alicia Menendez (University of Chicago)
Jack Thunde (AFLEARN)

AFLEARN
DATAHUB

CONTENTS

EXECUTIVE SUMMARY

Targeting teaching instruction by learning level rather than grade is an instructional approach that assesses students' current abilities, groups them accordingly, and tailors instruction to their actual learning levels. It aims to help children master foundational skills by meeting them where they are, rather than teaching strictly according to the grade-level curriculum. Originating in India and now widely implemented across sub-Saharan Africa, this approach has become known as Teaching at the Right Level (TaRL). The Global Education Evidence Advisory Panel (Akyeampong et al. 2023) has promoted this approach as a highly cost-effective "great buy." However, a closer reading of the evidence base reveals a more complex and context-dependent picture.

The Evidence Base

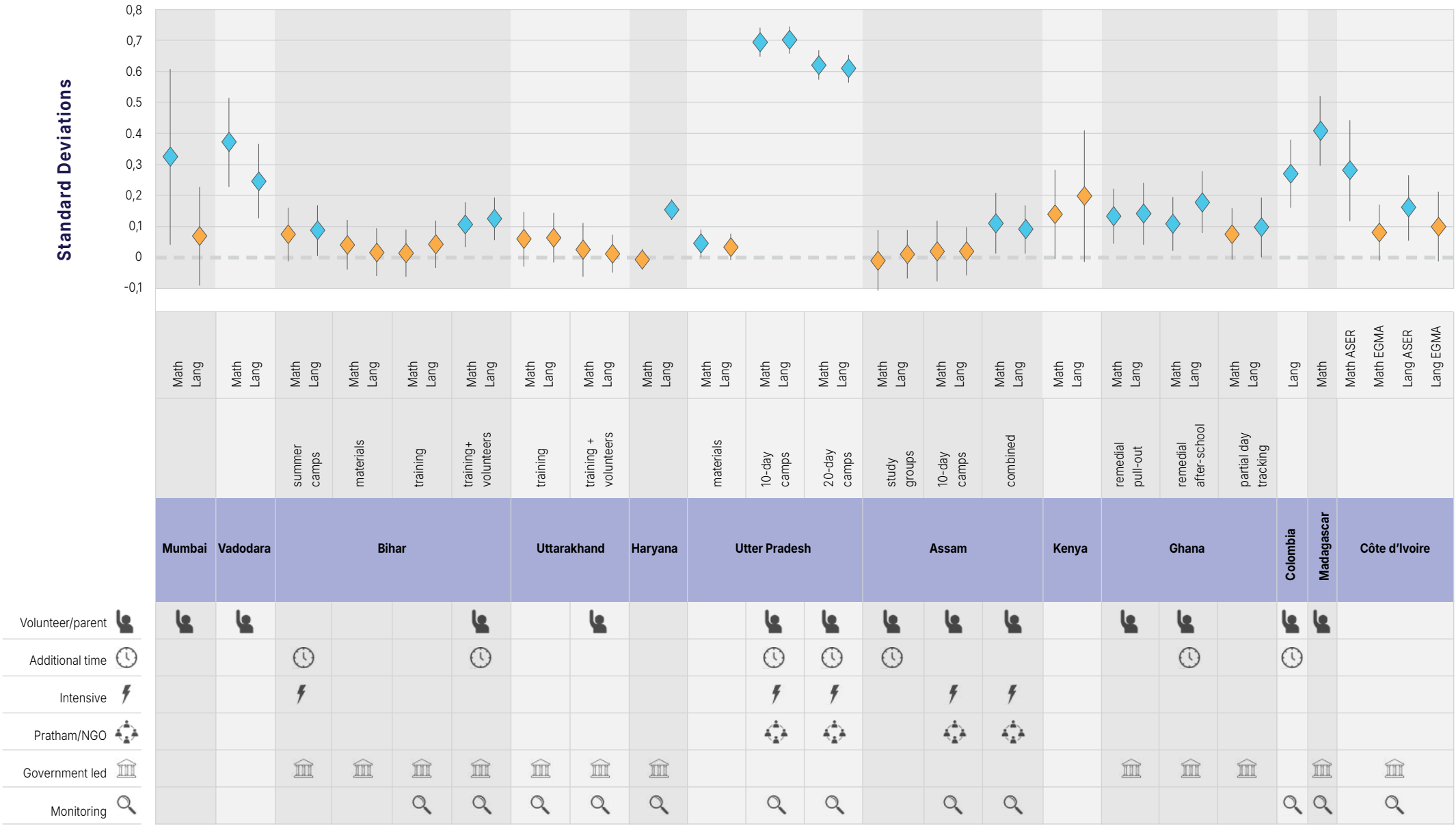
This critical review synthesises and evaluates the body of evidence on TaRL, drawing out key insights and providing an overarching assessment of its effectiveness. The aim is not to conduct an exhaustive systematic review, but rather to build upon the evidence cited in the three GEEAP reports (GEEAP 2020, 2022, Akyeampong et al. 2023), supplemented by more recent studies that meet the inclusion standards established by GEEAP. Specifically, eligible studies are those conducted in LMICs, employing a rigorous causal identification strategy, published as peer-reviewed journal articles or academic working papers and focusing on conventional TaRL models implemented at the primary school level.

Across experimental studies spanning India, Kenya, Ghana, Madagascar, Côte d'Ivoire, Colombia, the evidence is mixed. In total, there are 25 statistically significant results contrasted by 21 null effects. Statistically significant impacts range from 0.05 to 0.75 standard deviations (SD), but most fall below 0.2 SD.

The strongest results come from India, where TaRL originated. Repeated, short and intensive 10- or 20-day learning camps in Uttar Pradesh implemented by Pratham staff and volunteers produced large short-term gains of up to 0.7 SD. However, when a very similar model was replicated in Assam, no measurable impact was found, underscoring the difficulty of reproducing success even within the same country, organisation, materials and design. Further illustrating variation within the same programmeme, the Balsakhi tutoring programme achieved sizable improvements in literacy in Vadodara but no gains in Mumbai.

Outside India, results show similar heterogeneity. Recent evidence from Madagascar and Colombia found gains in volunteer- or tutor-led models while in Côte d'Ivoire, a teacher-implemented model produced only modest improvements on the ASER assessment and no effects on EGRA/EGMA. Likewise, Ghana's government partnership reported modest improvements (≈ 0.1–0.15 SD).

Figure 1. Estimated Effects of TaRL on Mathematics and Language (95% Confidence Intervals)



Statistically significant results are shown in blue while null results are indicated by orange.

Notes: 95% confidence intervals are based on normal approximations and may underestimate width, particularly for studies with fewer degrees of freedom.

Drivers of Variation

Variation in outcomes reflects differences in who delivers instruction, the amount of instructional time allocated, the intensity of coaching and monitoring, and the duration of implementation. In general, programmes implemented by NGO teams or trained facilitators under close supervision have achieved stronger results than those delivered by government teachers within existing school systems. Programmes that dedicate protected time each day for TaRL activities and receive consistent mentoring and oversight tend to show greater learning gains than those that expect teachers to incorporate TaRL informally into standard lessons without additional support. However, impacts often decline when interventions are scaled up or embedded in government systems, where maintaining fidelity, consistent training, and teacher motivation becomes more difficult. These patterns mirror the broader literature on implementation fidelity in education interventions (Banerjee et al. 2017; Angrist & Meager 2023).

Limits of Impact

Across contexts, TaRL gains are concentrated in basic skills such as letter recognition, decoding, number recognition and simple arithmetic with limited evidence of progression to higher-order comprehension or problem-solving. Few studies include post-intervention follow-ups; where they do, gains often fade or are substantially diminished within a year (Banerjee et al. 2007; Alvarez Marinelli et al. 2024).

Finally, in several studies (e.g., Ghana, Madagascar, Uttar Pradesh summer camps), TaRL added extra instructional time beyond the normal school day. Distinguishing between effects due to targeted pedagogy and those due to more time on task remains an open question.

Conclusion

Claims of a "strong body of evidence" are overstated. The balance of findings - mixed results from India, null effects from Kenya, modest gains from Ghana - does not support the GEEAP (Akyeampong et al. 2023) characterization of targeting teaching to learning levels and not grades as a universally "great buy."

Cumulative evidence indicates that TaRL can generate meaningful short-term improvements in foundational learning, particularly for the lowest-performing students, but its effectiveness is inconsistent and context-dependent.

The strongest gains have been observed under intensive, NGO-supported, short-term conditions that are difficult to replicate or institutionalize. As African countries pilot national scale-ups, forthcoming results from Zambia, Botswana, and Nigeria will be critical to determining whether TaRL can generate effectiveness within government systems.

1 INTRODUCTION

The global education landscape is facing a foundational learning crisis – especially in low- and middle-income countries (LMICs). Despite raising school enrollment, too many children are not mastering basic skills.

Nearly 60% of children worldwide cannot read and understand a simple text by age ten, with the figure rising to approximately 70% in LMICs (World Bank, 2022). These stark statistics represent more than just missed academic milestones – they carry long-term consequences for students' educational trajectories, economic prospects, and societal participation (Filmer et al., 2020).

Foundational learning – including basic literacy, numeracy, and core transferable skills – is essential for success in school and beyond. Without these building blocks, children are more likely to repeat grades, drop out early, and face reduced employment outcomes later in life (Glewwe et al., 2011; Hanushek & Woessmann, 2008). The impact is not evenly distributed: children from poorer or marginalized backgrounds are disproportionately affected, compounding existing inequalities within and between schools (UNESCO, 2021).

Several systemic factors contribute to this crisis. In many LMICs, national curricula are overloaded and misaligned with children's developmental levels. They often expect students to grasp complex content before mastering the basics (Pritchett & Beatty, 2015). Meanwhile, classroom instruction typically emphasizes rote memorization and procedural fluency at the expense of conceptual understanding (Hoadley, 2018). Teachers are frequently under-trained and lack the tools to assess individual learning gaps or provide differentiated instruction. Many have not received formal pedagogical training at all (Bold et al., 2019).

These challenges result in a persistent mismatch between what is taught and what students are actually ready to learn, perpetuating low learning levels and entrenching disadvantage from the earliest grades. One approach that has emerged to address this problem is targeting instruction to students' learning levels rather than their grades. Originating in India and subsequently spreading across sub-Saharan Africa, this approach has become known as Teaching at the Right Level (TaRL).

The Global Education Evidence Advisory Panel (GEEAP) identified interventions that tailor teaching to students' learning levels rather than their grade levels as a "great buy"—that is, interventions that are "highly cost-effective and supported by a strong body of evidence" (Akyeampong et al. 2023). However, the report offered only a highly summarized picture of the evidence, providing limited detail on variations in programmeme design, implementation, and contextual factors. Since then, a growing number of evaluations – particularly from sub-Saharan Africa - has expanded the evidence base and provided opportunities for an updated and more detailed analysis.

This critical review synthesises and evaluates the body of evidence on TaRL, drawing out key insights and providing an overarching assessment of its effectiveness. The aim is not to conduct an exhaustive systematic review, but rather to build upon the evidence cited in the three GEEAP reports (GEEAP 2020, 2022, Akyeampong et al. 2023), supplemented by more recent studies that meet the inclusion standards established by GEEAP. Specifically, eligible studies are those conducted in LMICs, employing a rigorous causal identification strategy, and published as peer-reviewed journal articles or academic working papers.

To qualify for inclusion, studies must examine programmemes that target instruction to students' current learning levels rather than their grade placement. The review focuses exclusively on conventional TaRL models, excluding technology-mediated interventions, and is restricted to programmemes implemented at the primary school level.

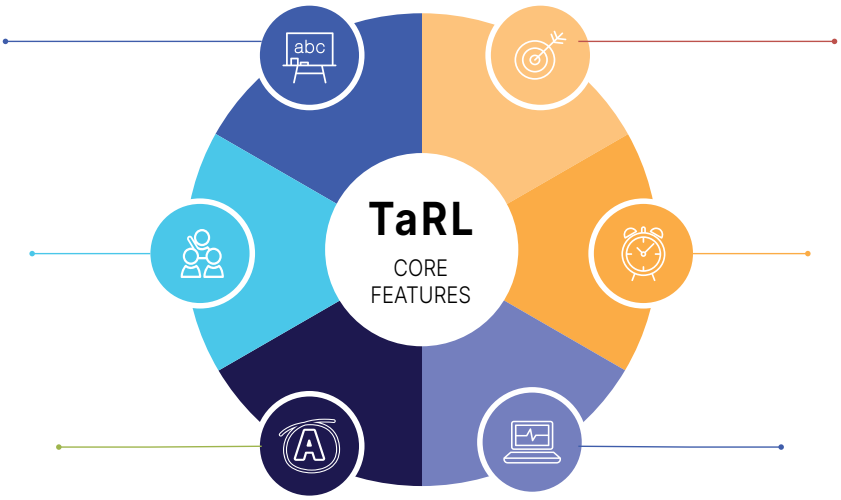
By bringing together this expanded body of evidence, the review seeks to provide a more nuanced understanding of the conditions under which TaRL is most effective. It begins by outlining the core principles of the TaRL approach and tracing its development in India and subsequent adaptation across Africa. The central section offers a critical synthesis of the evidence, highlighting both consistent patterns and key divergences across studies. By examining not only the impacts but also the implementation challenges and contextual factors that shape effectiveness, the review aims to provide a clearer picture of what this approach can—and cannot—contribute to addressing the global learning crisis.

2 CORE FEATURES OF THE TARL MODEL

Focus on Foundational Skills
Prioritises basic reading and arithmetic (rather than grade-level curriculum).

Grouping by Learning Level, Not Grade
Children are assessed and grouped according to current skill level.

Simple, Quick Assessments
One-on-one or small group tools to check basic reading and math ability.



Targeted Activities
Engaging, level-appropriate tasks designed for rapid progress.

Daily, Dedicated Time
Short bursts focused only on foundational skills. Conducted alongside, not instead of, the regular curriculum

Iterative Progress Monitoring
Frequent reassessment to regroup children as they improve.

TaRL is a learner-centered approach designed to address key drivers of low foundational learning in low- and middle-income countries (LMICs). Originally developed by Pratham in India and now implemented across several African countries through TaRL Africa and its partners, the approach aims to tackle two core challenges: wide learning disparities within classrooms and the rigid, age-based progression models that dominate many education systems (Banerjee et al., 2017; TaRL Africa, 2025a).

At its core, TaRL is built on a simple principle: start with what each child knows and build from there. Rather than following a fixed, grade-level curriculum, TaRL begins with a simple diagnostic assessment to determine each child's current reading or math level (Banerjee et al., 2016). Children are assessed using the ASER tool which was designed to capture basic learning outcomes using a single test administered to all children, regardless of age or grade. It is structured as a progressive assessment, with tasks organized by increasing levels of difficulty. Each child is evaluated according to the highest level they can successfully complete. In reading, the levels range from letter recognition to reading a Grade 2-level story, while in arithmetic, they span from number recognition to solving basic three-digit by one-digit division problems. Successful completion is defined at a basic threshold - for example, to be classified at the letter level, a child need only correctly identify any three letters from a set of ten. Once assessed, students are grouped by ability rather than age or grade and receive instruction that targets their actual learning needs. These groups are fluid, with learners regularly reassessed and regrouped as they progress.

Instruction under the TaRL approach is designed to be interactive and adapted to students' current learning levels. Activities often include group work, the use of familiar materials and methods such as storytelling and games to support engagement and understanding. The approach places emphasis on participation and peer interaction, which may differ from more traditional lecture-based classroom settings

TaRL focuses explicitly on foundational literacy and numeracy skills, recognizing that students cannot engage meaningfully with more advanced material without mastering the basics (World Bank, 2022). The TaRL approach can be implemented during school hours, in pull-out sessions, after school, or through short-term learning camps, depending on the context (TaRL Africa 2025a, Banerjee et al., 2017). Figure 1 illustrates the core components of the TaRL approach.

While the majority of studies included in this review evaluate comprehensive TaRL programmemes, several assess interventions that incorporate only selected elements of the TaRL approach - for example, a Kenyan study on tracking (Duflo et al. 2011), examined only the effects of assessing and grouping children by learning level. In other cases, TaRL forms part of a multi-component intervention, such as a programmeme in Madagascar (Maruyama and Igei 2024), which combined TaRL with school management strengthening activities.

3 EVOLUTION OF TARL AND INITIAL EVIDENCE BASE

TaRL originated from Pratham's efforts in India to address the growing recognition that school enrollment alone did not guarantee learning. Beginning in the early 2000s, data from Pratham's Annual Status of Education Report (ASER) consistently showed that many children in Grades 3–5 lacked basic literacy and numeracy skills. This learning crisis led Pratham to experiment with remedial education approaches aimed at helping children catch up to foundational skill levels.

These early initiatives laid the foundation for what would eventually become the structured TaRL methodology.

Table 1 summarizes the TaRL implementation models that emerged over the period 2001 to 2019, outlining their delivery mechanisms, implementing entities, and allocation of time in and out of school. Appendix Table A1 provides a summary of the accompanying evaluations, detailing sample size, outcome measures, duration and grades.

Table 1: Evolution of TaRL Implementation Models in India

Location	Years	Intervention	Implementer	Coaching / monitoring	School Time	Additional Time	Duration	Grades	Government involvement	Key reference(s)
Mumbai, Maharashtra	2001 - 2003	Balsakhi	Trained volunteers	NO	120 min per day	NO	24 months	3 and 4	NO	Banerjee et al. (2007)
Vadodara, Gujarat										
Uttar Pradesh	2005 - 2006	reading camps	Trained volunteers	YES, periodic visits from Pratham	NO	YES, after school	12 months	ages 7-14	NO	Banerjee et al. (2010)
Bihar	2008	summer camps	Trained and paid regular teachers + trained volunteers	NO	NO	YES, summer	1 month	1 to 5	YES, evaluation embedded in full scale-up to 40,000 schools	Banerjee et al. (2016); Banerjee et al. (2017)
	2008 - 2010	materials only	Regular teachers	NO	YES, unspecified	NO	24 months	1 to 5		
		training	Trained regular teachers	YES, monitoring by Pratham		NO				
		training + volunteers	Trained regular teachers + trained volunteers outside school hours			YES, after school				
Uttarakhand	2008 - 2010	training	Trained regular teachers	YES, monitoring by Pratham	YES, unspecified	NO	24 months	1 to 5	YES, evaluation embedded in full-scale up to 12,000 schools	
		training + volunteers	Trained regular teachers + trained volunteers during school hours							
Haryana	2012 - 2013	dedicated hour in school day	Trained (by govt officials trained by Pratham) regular teachers	YES through trained block coordinators	60 min per day	NO	12 months	3 to 5	YES, programme fully supported and implemented by government	
Uttar Pradesh	2013 - 2014	materials only	Regular teachers	NO	YES, unspecified	NO	12 months	3 to 5	NO	
		four 10-day camps	Pratham team members assisted by trained volunteers	YES, conducted by Pratham	180 min per day	YES, summer				
		two 20-day camps								
Assam	2018 - 2019	study group	Managed by parents	NO	NO	YES, after school	16 months	1 to 5	NO	Nyqvist and Guariso (2022)
		three 10-day camps	Pratham team members with the support of teachers and volunteers.	YES, conducted by Pratham	YES, unspecified	NO				

Early Remedial Models and First RCTs

One of the earliest evaluations of remedial instruction focused on the "Balsakhi" Remedial Education Programme (Banerjee et al. 2007). Starting in 2001, this intervention recruited and trained young women from local communities to serve as "Balsakhis" (or "friends of the child"). These Balsakhis provided remedial instruction in basic literacy and numeracy to children in Grades 3 and 4 who were falling behind their peers. The instruction occurred during school hours, with selected children pulled out of their regular classrooms for two hours each day to receive targeted lessons in small groups, using a standardized curriculum developed by Pratham. The goal was to provide individualized support that better matched each student's actual learning level.

The study used randomized controlled trials across 77 and 122 schools in Mumbai and Vadodara respectively. The impact varied across locations. In Mumbai, the Balsakhi programme improved math scores by 0.16 SD in Year 1 and 0.32 SD in Year 2, but for literacy, changes in test scores were not statistically significant. In Vadodara, the impact on math scores was 0.19 SD in Year 1 and 0.37 SD in Year 2, while literacy scores improved by 0.11 SD and 0.25 SD in Years 1 and 2, respectively.

While the evaluation showed significant short-term gains, the positive effects faded for most students one year after the programme ended (Banerjee et al. 2007). The average treatment effect was no longer statistically significant, and although gains among the lowest-performing students persisted, they were reduced in magnitude.

Remedial Camps

After-school remedial reading camps in Utter Pradesh in 2005 were evaluated by Banerjee et al. (2010). In this intervention, local youth volunteers were trained in TaRL pedagogical methods and supported by Pratham to run after-school remedial reading camps in their villages. Children were assessed and grouped by reading level, and instruction focused on helping them progress step-by-step through the reading continuum using interactive and game-based activities.

Student participation was low, with only 8 percent of children attending the camps. As a result, the intention-to-treat (ITT) estimates are modest with a 1.7 percentage point increase in the proportion of children classified at the letter level, 1.8 percentage point increase in the proportion of children classified at the word and paragraph level combined and no increase in the proportion of children classified at the story level[1].

The next iteration of remedial camp involved government teachers in Bihar during the summer of 2008 (Banerjee et al. 2016, 2017). During this one-month programme government teachers, trained and paid for summer service, and trained volunteers used Pratham materials to provide remedial instruction to children in grades 1–5. The programme had 23% attendance among eligible children. The statistically significant ITT estimated gains were 0.09 SD in language and 0.07 SD in math.

The subsequent remedial camps evaluated by Banerjee et al. (2017) were conducted during school hours in Uttar Pradesh in 2013/14. These intensive TaRL-style learning camps temporarily replaced regular classroom instruction.

The study included four groups:

1. A group that received 10-day learning camps repeated over four cycles, plus a 10-day summer camp.
2. A group that received two cycles of 20-day camps, lus a 10-day summer camp.
3. A materials-only group that received TaRL resources without training or structured support.
4. A control group that continued with standard classroom instruction.

[1] Using an instrumental variable (IV) approach, the study finds a 22 percentage point increase in the proportion of children able to read at least letters, a 23 percentage point increase for those reading at least a word or paragraph, and a 22 percentage point increase in those able to read stories, though this last effect is not statistically significant.

The camps were led by Pratham instructors, supported by supervised volunteers, and conducted during school hours, suspending normal lessons. Students in grades 3 to 5 were grouped by ability and taught Hindi and Math for approximately 1.5 hours each day using games and level-appropriate materials. Both the 10-day and 20-day camp models led to large improvements in reading and math: the 10-day model increased scores by about 0.7 SD in both subjects, while the 20-day model raised reading scores by 0.61 SD and math scores by 0.62 SD. These effects correspond to an average increase of between 0.73 to 0.95 ASER levels. For reading, children were 16 percentage points less likely to be classified at the beginner level (unable to identify letters) and 25 percentage points more likely to be classified at the paragraph or story level. The materials-only group showed small but statistically significant gains of 0.045 SD in math and no gains in language.

Nyqvist and Guariso (2022) evaluated a similar reading camp model implemented by Pratham in the state of Assam between 2018–2019. The intervention combined in-school reading camps with an out-of-school community study group component managed by parent volunteers. The study randomly assigned villages to four groups: - a control group and three intervention arms:

1. Three cycles of ten-day reading camps delivered over five months
2. Community-based study groups only
3. Both interventions combined

Following the model previously tested in Uttar Pradesh, the learning camps were conducted by Pratham teams during regular school hours. Pratham introduced the camps at the start of the school term, mobilizing community volunteers to support implementation. Camps were administered by Pratham staff, with assistance from available teachers and volunteers.

The study groups were introduced by Pratham in target villages, mobilizing community volunteers to manage them. Groups comprised around six primary-school children and were coordinated by community volunteers, typically mothers. The number of groups per village varied based on the availability of children and volunteers. Group structures were deliberately flexible: volunteers decided how often and how long to meet. While no formal records were kept, a qualitative study of 40 volunteers revealed that most groups met two or three times per week for about two hours. Pratham supplied self-explanatory learning materials designed to guide activities, but volunteers were also encouraged to support children with homework and reading practice.

In this setting, neither the camps nor the study groups alone produced significant learning gains. However, the combined intervention led to modest but statistically significant improvements of 0.11 SD in literacy and 0.09 SD in mathematics.

These findings contrast sharply with results from Uttar Pradesh, where the same pedagogical model, implementing partner, data collection agency and assessment tools produced large gains (0.58 SD in mathematics and 0.69 SD in reading; Banerjee et al., 2016). Using combined data from both studies, Nyqvist and Guariso ruled out several potential explanations—such as baseline learning levels, student demographics, survey timing, and school characteristics—but could not pinpoint the factors driving the difference. This persistent gap underscores the challenges of replicating successful programmes across contexts and time.

TaRL in Government Systems

Banerjee et al. (2017) evaluated a series of Teaching at the Right Level (TaRL) programmemes integrated into government systems across three Indian states. In Bihar and Uttarakhand, the evaluation was embedded within large-scale government rollouts that reached the majority of schools in each state. Between 2008 and 2010, three distinct implementation models were tested across treatment villages:

1. Materials-only (M): Distribution of Pratham learning materials during the school year without training or additional support.
2. Teachers and Materials (TM): Distribution of Pratham learning materials during the school year, teacher training and monitoring support.
3. Teachers, Materials, and Volunteers (TMV): The most intensive model, adding community volunteers to assist struggling students outside school hours.

In Bihar, only the TMV model produced statistically significant improvements—0.13 SD or 0.2 ASER levels in language and 0.11 SD or 0.13 ASER levels in math—suggesting that the involvement of volunteers outside regular school hours was crucial. The M and TM interventions did not yield significant effects.

In Uttarakhand, where volunteers worked during school hours to support teachers rather than operating separately, neither the TM nor TMV interventions led to significant gains. The authors attribute this lack of impact to weak implementation fidelity and the absence of structured time devoted to TaRL, despite positive reception of

the training and ongoing Pratham support.

The next intervention, in the state of Haryana, attempted to more firmly institutionalize the TaRL approach with the government system. The programmeme was explicitly framed as a government-led effort, supported and implemented by the Department of Education. To reinforce this perception, a new cadre of academic leaders, known as Associate Block Resource Coordinators (ABRCs) was established to train and supervise teachers during implementation.

Pratham provided the ABRCs with four days of training and field practice, after which they spent 15–20 days in schools conducting daily classes and testing the methodology of grouping students by learning level and delivering level-appropriate instruction. Once this practice phase concluded, the trained ABRCs—supported by Pratham staff—went on to train teachers within their assigned jurisdictions.

A key design feature was the introduction of a dedicated instructional hour within the official school timetable during which students in Grades 3–5 were reassigned from grade-based classrooms into achievement-based groups, based on a baseline assessment conducted jointly by teachers and coordinators.

The intervention and evaluation were conducted in 2012-2013 with 200 of the 400 schools assigned to the TaRL intervention2. Initially, schools focused on both Hindi and Math, but midline assessments revealed limited progress. As a result, Pratham advised schools to drop Math and focus solely on Hindi (Duflo et al. 2015). The evaluation found significant gains in Hindi (0.15SD or 0.2 ASER levels) but no improvements in Math.

4 EVIDENCE OF TARL IN OTHER CONTEXTS

Following the early evidence from India, the TaRL approach was introduced and adapted in a range of African low- and middle-income countries (LMICs) over the past decade. These efforts were typically implemented through partnerships involving Pratham (or TaRL Africa), national governments, NGOs, and research institutions. Evaluations using randomized controlled trials and quasi-experimental methods have generated and continue to generate a growing body of evidence on the effectiveness of TaRL-style interventions in diverse African settings. This section synthesises findings, with particular attention to intervention designs, implementation challenges, and contextual factors.

Table 2 provides a summary of the TaRL-related interventions outside of India. Details of the accompanying evaluations can be found in Appendix Table 1.

[2] The evaluation included both the TaRL-based intervention, known as the Learning Enhancement Programmeme (LEP), and the Continuous and Comprehensive Evaluation (CCE) that replaced traditional examinations with frequent teacher-led assessments aimed at improving classroom feedback. The study randomly assigned 400 primary schools to one of four groups: LEP alone, CCE alone, both LEP and CCE, or neither (control).

Table 2. TaRL interventions outside of India

Location	Years	Intervention	Implementer	Coaching / monitoring	School Time	Additional Time	Duration	Grades	Government involvement	Key reference
Kenya	2005 - 2006	tracking	Teacher and contract teacher	NO	Full day	NO	18 months	1		Duflo et al. (2011)
Ghana	2010 - 2013	remedial pull-out	Gov. paid youth assistant	NO	YES, 60 min per day x 4 days per week	NO	36 months	1 to 3	Implemented in partnership with education ministry, teacher assistants from government youth employment programmeme	Duflo et al. (2024)
		remedial after-school	Gov. paid youth assistant		NO	YES, 60 min per day x 4 days per week				
		partial day tracking	Teacher		YES, 60 min per day x 4 days per week	NO				
Madagascar	2019	TaRL plus school management	Teacher and community volunteer	YES	NO	YES	3-4 months	3 to 5	Developed by education ministry	Maruyama and Igei (2024)
Colombia	2015 - 2017	remedial education	Trained tutors, (univ. students/ recent graduates)	YES	YES, 40 min per day	NO	3-4 months	3		Alvarez Marinelli et al. (2024)
Côte d'Ivoire	2021 - 2023	TaRL	Trained teachers	YES, through trained government district-level pedagogical advisors + Chatbot	YES, 45 min per day	NO	18 months	3 to 5	Implemented by ministry of education in collaboration with TaRL Africa	Wolf et al. (2025)

Kenya: Evidence from Tracking by Prior Achievement

While not a direct implementation of TaRL, Duflo et al. (2011) conducted a randomized evaluation of ability-based tracking in an African context to determine whether grouping students by ability could positively influence learning outcomes The study targeted first grade students in 121 primary schools, each of which initially had a single first-grade classroom. To enable tracking, a contract teacher was hired in each school, and students were split into two first-grade classes. One class was taught by a civil service teacher, the other by the contract teacher, with the pairing randomized. Random student placement was maintained in 61 randomly selected schools, while students were grouped by prior achievement in the remaining 60 schools.

Students were assessed using standardized math and language tests, with components designed by a cognitive psychologist to capture a broad range of skills expected by the end of grade 2. The assessment included both written and orally administered one-on-one components, evaluating abilities from letter and number recognition to three-digit subtraction and sentence comprehension.

After 18 months, students in tracked classrooms showed modest improvements, with average test score gains of 0.14 SD. These effects persisted one year later at 0.16 SD. However, the results were not statistically significant at conventional levels. After controlling for baseline scores, students in the top half of the pre-assignment distribution gained 0.19 SD, and those in the bottom half gained 0.16 SD. Students in the middle of the distribution benefited as much as those at the top and bottom. Interestingly, the lowest-scoring student assigned to the higher-achieving section and the highest-scoring student assigned to the low-achievement section performed similarly by endline, suggesting that peer group composition did not significantly affect individual learning under tracking. Additionally, the study found that students in the lower half of the initial distribution gained more from tracking in basic math skills, while those in the upper half gained more in advanced math skills[3].

Although the results were not significant at conventional levels, these findings have been widely interpreted as early evidence that grouping students by ability can positively influence learning outcomes, even outside a formal TaRL model.

Ghana: Testing Alternative Delivery Models at Scale

In Ghana, Duflo et al. (2022) conducted a multi-arm randomized controlled trial to evaluate different models of TaRL-style remedial instruction. The study was implemented in partnership with the Ghana Education Service (GES) and focused on interventions targeting students in lower primary grades (grades 1, 2, and 3) across government schools for three years. The aim was to assess whether different delivery approaches to remedial education could improve foundational literacy and numeracy outcomes.

The evaluation tested four interventions, three of which closely aligned with the TaRL methodology:

1. Pull-out remedial: Teaching assistants worked with remedial learners on a pull-out basis during the school day. Educators tested students at the start of each term to determine their learning level and grouped remedial learners into smaller, more homogenous classrooms for part of the day.

2. After-school remedial: Teaching assistants worked with remedial learners outside of the school day. Educators tested students at the start of each term to determine their learning level and remedial learners received an extra instructional hour.

3. Classroom split: Teaching assistants worked with half of the classroom each day on grade-level content. This reduced classroom size for all students resulting in reduced teacher-to-learner ratios.

4. Partial day tracking: Trained teachers divided students into three learning levels for part of the day and focused instruction on students' learning levels. Educators tested students at the start of each term to determine their learning levels.

[3] Basic maths refers to addition or subtraction of single-digit numbers. Advanced maths refers to addition or subtraction of three-digit numbers.

The teacher assistants were provided and paid through the National Youth Employment Programme (NYEP) under the Ministry of Youth and Sports.

Student learning was assessed using standardized tests in math, English, and local language. The assessments were developed in collaboration with the Assessment Services Unit of the Curriculum Research and Development Division of GED. They included both foundational and grade-level content aligned to the official curriculum for grades 1 to 3 – extending beyond the content typically found in ASER, EGRA, or EGMA tools.

For the three models that aligned with TaRL, learning gains were observed although effects varied:

- Pull-out remedial: Increased overall learning by 0.11 SD in Year 2 and 0.14 SD in Year 3. On foundational questions alone, gains were 0.14 SD, and 0.15 SD.
- After-school remedial: Overall gains of 0.11 SD and 0.15 SD in Years 2 and 3; 0.14 SD and 0.15 SD on foundational questions.
- Partial-day tracking: No overall gains in Year 2 and gains of 0.08 SD in Year 3; however, gains of 0.11 SD and 0.13 SD in Years 2 and 3 respectively for foundational questions. No gains persisted post-intervention.

These estimated gains of these three models are not significantly different from each other unless only foundational questions are considered. When focusing on foundational questions the two remedial interventions have larger and statistically different impacts than the tracking model.

In general girls improved more than boys by around 0.10-0.15 SD. While the interventions improved foundational skills, it remains unclear whether they also led to improvements in more advanced, grade-level content. Since analyses either examined all questions collectively or isolated foundational content, it is difficult to determine the extent to which gains reflect broader curriculum mastery.

Implementation challenges were noted across all models. Teachers consistently taught their assigned groups in only about one-third of observed sessions, partly due to delays in material provision, payment issues and absenteeism. These findings suggest that while multiple models of level-based instruction have potential for effectiveness, fidelity of implementation is a key constraint in real-world settings.

Colombia: Evidence from a Sequence of Experiments Colombia

An intervention focused on improving the reading skills of struggling third grade students in Columbia was evaluated by Alvarez Marinelli et al. (2024). The intervention included three key elements – content targeted at the right level, phonics-based methods and highly structured materials – and incorporates core TaRL features such as targeting foundational literacy skills and working in small groups based on assessed needs.

Structured 40-minute remedial tutorial sessions (20 minutes for reading fluency related exercises, ten minutes for vocabulary building, and the other ten for reading comprehension strategies) were conducted three times a week during school hours. Groups were small, with no more than six students.

The sessions were held during school hours, specifically during the second half of the academic year (after the mid-year break in June). The duration of the intervention was 12 weeks in the first cohort and 16 weeks in the next two cohorts.

Tutors, mainly university students or recent graduates, were trained to implement the sessions. The training included an eight-hour initial session, regular coaching meetings, and two supervised on-site sessions.

The evaluation used a randomized experimental design at the school level. In the first cohort, 94 schools participated. In the second and third cohorts, the sample size was reduced to 84 and 80 schools respectively. Assessments were conducted using EGRA administered by trained assessors.

The evaluation study found overall gains of 0.27 SD in literacy scores at the end of third grade compared to the control group. These gains were driven by a 0.35 SD increase in letter sound knowledge (4 additional letter sounds) and a 0.16 SD increase in oral reading fluency (2 additional words). There was no impact on reading comprehension. The impact persisted by the end of grade 4 but with a 40 percent reduction. The gains by the end of the third grade on the aggregate literacy scores increased as the programme went on, growing from 0.14 in Cohort 1, to 0.22 in Cohort 2, to 0.55 in Cohort 3. Researchers attributed these changes to improvements in the programme and longer interventions.

Madagascar: Community-Driven Remedial Support Anchored in TaRL Principles

Maruyama and Igei (2024) evaluated a randomized implementation of PMAQ-TaRL, a community-anchored package of interventions developed by Madagascar's Ministry of Education with technical support from Japan International Cooperation Agency (JICA). The programme combined a structured school management strengthening component with the TaRL instructional approach and targeted students in grades 3 to 5 in the Amoron'i Mania region.

The intervention package was rolled out across 70 public primary schools. It featured 1) the establishment and capacity-building of school management committees (SMCs), 2) collaborative development of school action plans incorporating extracurricular remedial activities, and 3) training of teachers and community volunteers in TaRL pedagogy and math workbooks together with student training on how to use the workbooks. The TaRL component focused on foundational skills in basic reading and mathematics (numbers and the four operations), delivered through extracurricular remedial classes based on students' assessed learning levels. These classes were primarily organized by teachers and totaled an average of 80 hours per school over the 3.5 month period of the intervention. In the majority (84%) of school, remedial classes were supported by at least one community volunteer.

The results showed a significant improvement in foundational numeracy, with average gains of 0.41 SD in test scores across grades 3 to 5. The largest effects were observed among grade 3 students (0.47 SD).

The effects were stronger for lower-achieving students, suggesting that the intervention was particularly effective at reaching children who were furthest behind. In reading, the proportion of students reaching the paragraph and story levels increased by 3.4 and 18.4 percentage points, respectively, although gains were smaller for students who started at the beginner level.

Unfortunately, it is not possible to identify the contribution of the different components included in this intervention.

Côte d'Ivoire: Programmeme d'Enseignement Ciblé (PEC)

In Côte d'Ivoire, Wolf et al. (2025) evaluated the Programmeme d'Enseignement Ciblé (PEC), implemented by the Ministry of National Education and Literacy (MENA) in collaboration with TaRL Africa. The RCT covered 167 schools, 303 teachers, and 3808 students across 3 regions from fall 2021 through spring 2023. The intervention embedded within the public education system, involved daily 45-minute sessions on targeted small-group instruction in foundational French literacy and mathematics.

Teachers conducted baseline assessments to group students by proficiency levels, delivered playful and level-appropriate instruction across grades 3, 4, and 5, and reassessed students at mid-year and end-year to track progress and regroup as needed. The implementation was supported by trained government mentors who provided pedagogical guidance throughout the intervention.

The evaluation reported significant improvements in foundational learning levels as measured by ASER, with effect sizes of 0.16 SD in literacy and 0.28 SD in numeracy. This corresponds to a statistically significant 7 percentage point (19% in control versus 11% in treatment) reduction in children in the beginner level for literacy. For numeracy, there was a 10 percentage point reduction in children classified at the two-digit level (level 2) and a 9 percentage point increase in children classified at the subtraction level (level 4). However, impacts measured by the EGRA and EGMA were not significant. The researchers' interpretation of these findings points to a disconnect between foundational learning classifications and actual functional performance. As examples, they mention that students categorized at the subtraction level on ASER scored only 55% on the EGMA (sequencing, addition, subtraction composite) and 35% on the subtraction subtest. Similarly, students classified as story-level readers on ASER scored only 50% on the EGRA (letter, word, nonword reading composite).

5 ONGOING EVALUATIONS OF GOVERNMENT-LED SCALE-UPS ACROSS AFRICA

Government-led scale-ups of TaRL-based initiatives are underway in Zambia, Nigeria and Botswana with associated evaluations planned or in progress.

Zambia

Zambia's Ministry of General Education, with support from Pratham, J-PAL Africa, UNICEF, and USAID, piloted the Catch-Up programme—a TaRL-based remedial initiative—between 2016 and 2018. Adapted from India's TaRL model for Zambia's multilingual context, it targeted grades 3–5 across several provinces (Lipovsek et al. 2023). Teachers were trained to assess students, group them by ability, and use level-appropriate activities for one hour daily, with ongoing mentoring from government officials and NGO staff (Innovations for Poverty Action, 2017). Following an 80-school pilot in 2016, the programme scaled to about 1,800 schools by 2018. An RCT across 273 schools is underway, with findings expected to shed light on the effectiveness of TaRL when it is institutionalized and implemented at scale through government systems (de Barros et al. 2023).

Botswana

In Botswana, the Ministry of Education is implementing a national scale-up of a TaRL programme in partnership with the NGO, Youth Impact. By 2022, the intervention had reached approximately 20% of government primary schools, with plans for national coverage by 2026 (Angrist and Meager 2023). The intervention targets children in grades 3 to 5 for one hour per day for a six-week period. The programme includes youth participating in the government's National Service Programme (NSP) (TaRL Africa 2025b). Youth Impact are currently conducting a quasi-experimental evaluation of the programme using a difference-in-differences approach.

Nigeria

In Nigeria, TaRL has been scaled up across five states (Angrist and Meager 2023). Government teachers deliver TaRL classes for 1.5 to 2 hours within each school day to children in grades 4 to 6. Support is provided by cluster-level government support officers and head teachers within the same school (TaRL Africa 2025c). A 150 school RCT is planned for this school year[4].

[4] *Email correspondence with TaRL Africa.*

6 WHAT THE EVIDENCE TELLS US – LEARNING GAINS AND PATTERNS:

The emerging evidence highlights both the potential and challenges of implementing TaRL approaches.

While some programmemes have achieved encouraging results in improving foundational skills, their effectiveness depends critically on factors such as take-up, implementation quality, alignment with existing systems, and the level of ongoing support. This section synthesises findings from the available evaluations, identifies key patterns, and offers a nuanced interpretation of the results.

Learning Gains in RCTs

Figure 2 presents a visual summary of impacts across studies, showing 95 percent confidence intervals and distinguishing null and statistically significant results orange and blue respectively[5]. Although significant effect sizes range from 0.045 to 0.745 standard deviations, the majority are small to modest (below 0.2). Overall, the evidence base is mixed, with 25 statistically significant results contrasted by 21 null effects[6]. Excluding the intervention arms that only included materials, there are 24 statistically significant and 18 null effects. Even among interventions following similar implementation models, results vary considerably. For instance, the largest gains – by a substantial margin – were achieved in the 10- and 20-day reading camps in Uttar Pradesh, yet a replication of the camp-based design - implemented by the same organisation using the same materials and pedagogy – had no measurable impact in Assam. Similarly, the Balsakhi programmeme generated heterogeneous results across sites, with no statistically significant effect on literacy in Mumbai despite positive findings in Vadodara.

These discrepancies highlight that TaRL's effectiveness is context-dependent.

Across the literature, effect sizes vary substantially by context and delivery model, reflecting differences in who delivers instruction, how much time is allocated, the intensity of monitoring and coaching, and overall duration of the intervention. In general, models implemented by NGO teams or trained facilitators under close supervision have produced stronger results than those delivered by government teachers within existing school systems. Similarly, programmes that allocate dedicated time for TaRL activities and receive regular mentoring and monitoring tend to achieve higher gains than those that expect teachers to integrate TaRL into normal classroom routines without additional support.

Impacts also tend to decline when programmes are implemented at larger scale or through government systems, where fidelity and teacher support are harder to sustain. In India, for example, the learning camps and Pratham-trained assistant models generated some of the largest improvements, while government-led versions without structured support produced modest or negligible effects.

Outside India, results show similar heterogeneity. Recent evidence from Madagascar and Colombia found gains in models involving community volunteers or tutors, but in Côte d'Ivoire, a teacher-implemented model produced only modest improvements on the ASER assessment and no effects on EGRA and EGMA.

[5] *The figure does not include language estimates for the 2005 reading camp in Uttar Pradesh and the intervention in Madagascar, as effects were not reported in standard deviations, rather they estimate the impact of treatment on the probability of being at each of the ASER level. In Utter Pradesh, the camps lead to a 1.7 percentage point increase in the proportion of children classified at the letter level, 1.8 percentage point increase in the proportion of children classified at the word and paragraph level combined and no increase in the proportion of children classified at the story level children. In Madagascar, across all grades children are 18.4 and 3.4 percentage points more likely to be at the story and paragraph levels respectively. These effects are significant at the 1% and 5% respectively (Maruyanma and Igei 2024).*

[6] *These numbers include the literacy outcomes for the 2005 camps in Uttar Pradesh and the Madagascar intervention.*

Likewise, TaRL programmemes implemented through government partnerships in Ghana yielded positive but modest effects. Ongoing studies in Zambia, Botswana, and Nigeria will be critical to determining whether TaRL can achieve consistent results when embedded within government systems.

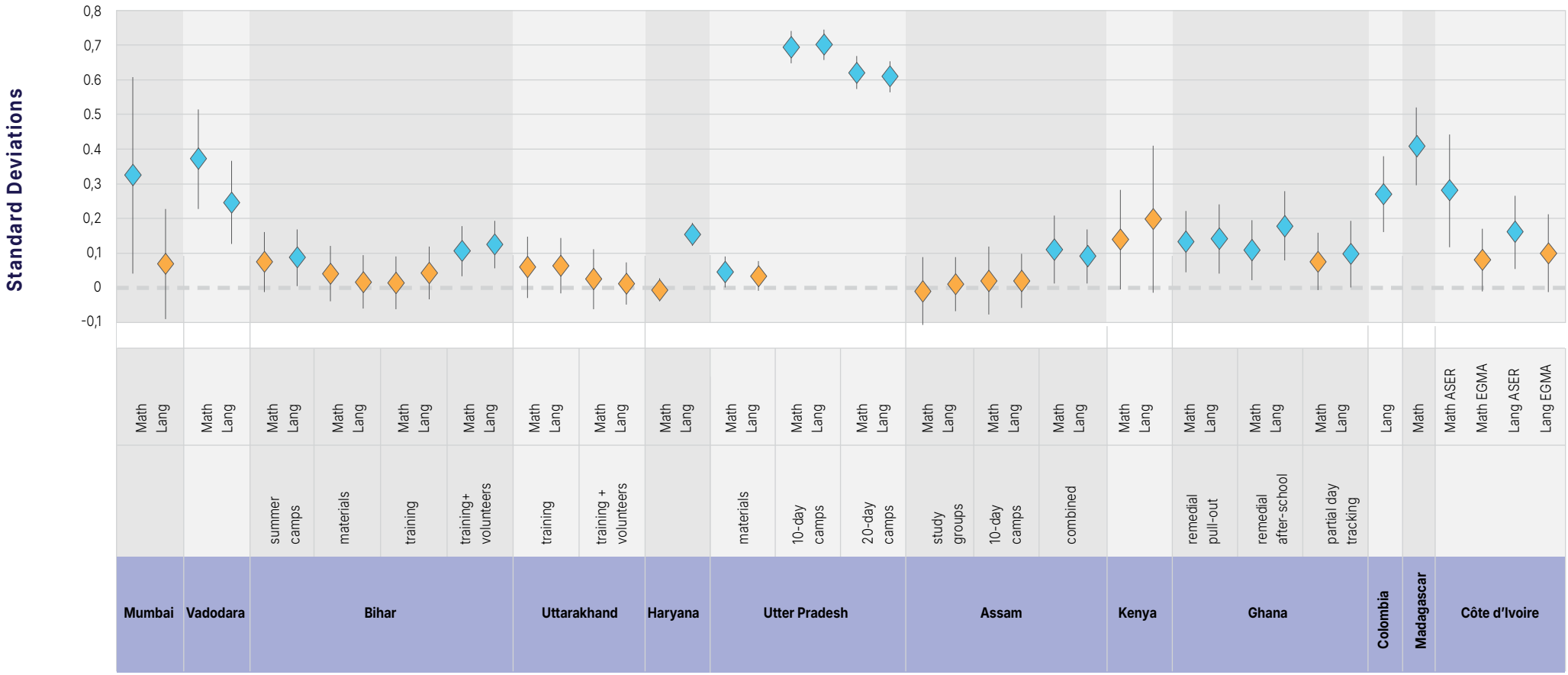
In addition to implementation fidelity, low participation often undermines impact. In 2005 in Uttar Pradesh, for example, only 8 percent of eligible students attended reading camps. Angrist and Meager (2023) documented wide variation in take-up (8 to 90 percent) and fidelity (23 to 83 percent) across Indian and Kenyan programmes. They argue that treatment-on-treated (TOT) estimates can offer insight into potential effectiveness under ideal conditions, yet these estimates risk conflating programme participation with unobserved factors such as teacher motivation or school accountability. For policy purposes, intention-to-treat (ITT) estimates remain more appropriate, as they reflect real-world implementation constraints.

It is instructive to consider the African programmeme designs in light of the evidence from India. The government-led scale-ups in Zambia, Côte d'Ivoire, and Nigeria most closely resemble India's Haryana study, which produced only modest literacy gains and eventually dropped mathematics due to implementation challenges. In Botswana, elements of the volunteer learning-camp model have been incorporated through the government's National Service Programmeme, yet the dosage of one hour per day for six weeks falls substantially short of the intensive Uttar Pradesh and Assam learning-camp models.

The 2023 GEEAP report identified interventions that target instruction to students' learning levels as a "great buy" based on evidence from the first Balsakhi programme in India through to the study in Ghana. However, when viewed collectively, the body of evidence does not appear to justify such a strong endorsement. The pattern of findings—mixed results from India, null effects from Kenya, and only modest improvements from Ghana—falls short of constituting the "strong body of evidence" that the GEEAP designation implies.

Taken together, these studies suggest that while TaRL can produce learning gains under certain conditions, its effectiveness is highly context-dependent and far less consistent than the "great buy" label would suggest.

Figure 2. Estimated effects on Mathematics and Language – 95% confidence intervals



Statistically significant results are shown in blue while null results are indicated by orange.

Notes: 95% confidence intervals are based on normal approximations and may underestimate width, particularly for studies with fewer degrees of freedom.

Standardized outcomes measures

Standardized effect sizes, while commonly reported, are a problematic metric for interpreting the impact of education interventions.

By construction, standard deviations depend on the underlying distribution of test scores, which varies across populations, instruments, and methods of score aggregation (Coe 2002, Simpson 2017). As a result, comparisons across studies or contexts can be misleading.

The issue is particularly acute when ordinal outcomes with a small number of discrete categories—such as the ASER reading levels—are converted into standardized units. Such transformations assume the data are normally distributed with equal intervals between levels. However, the step from one ASER level to the next represents a non-linear and often substantial leap in skill. In the presence of floor ceiling effects, the data will be highly skewed and the distribution compressed. When such data are transformed into standard deviations, the resulting estimates can distort both the magnitude and meaning of the effects. For example, in contexts where most students cluster at the lowest ASER levels, even modest shifts upward can yield disproportionately large standardized effects.

For continuous measures, such as fluency scores from EGRA, standardization can also obscure meaning, since the aggregation of subtasks (e.g., letter identification, word reading, comprehension) and differences in score scaling can strongly influence the resulting standard deviation.

Additionally, standardized gains expressed in SD units often fail to capture whether the learning improvements are educationally or pedagogically meaningful (Stern and Piper 2019). For example, a reported gain of 0.35 SDs in letter-sound knowledge in Colombia translated to students learning only four additional letter sounds—statistically significant but limited in practical terms. Expressing outcomes in their original assessment units provides clearer insights. In Côte d'Ivoire, a statistically significant effect of 0.16 SDs on ASER reading corresponded to only a one percentage-point increase in the proportion of students reaching the paragraph or story level. In contrast, Madagascar recorded increases of 3.4 percentage points at the paragraph level and 18.4 percentage points at the story level. These comparisons underscore the importance of reporting effects in absolute and interpretable terms, particularly when dealing with ordinal or bounded learning measures such as ASER.

Limited Gains Beyond Foundational Skills

Evidence suggests that TaRL's learning gains are concentrated in the most basic competencies. In Côte d'Ivoire, significant gains were observed on the basic ASER assessment, but no impact was found on the more demanding but still foundational EGRA and EGMA (for example, the EGRA assessment only included letter-reading, word-reading, and pseudoword-reading and did not test higher order skills such as comprehension). In Colombia, the largest gains were in letter-sound knowledge, with no effect on reading comprehension. Similarly, in Ghana, impacts were stronger for foundational skills, though results did not clearly indicate whether curriculum-aligned skills improved. This raises questions about the ceiling of TaRL's effectiveness, especially for students approaching grade-level content. If the model does not support progression to more complex skills, its role in long-term educational advancement may be limited.

Test Familiarity

An additional concern is potential bias in studies that use the ASER assessment as the outcome measure. Employing the same tool that is used to form ability groups throughout the intervention, creates the risk of better performance in the treatment group due to test familiarity. This may explain some of the divergence between ASER and EGRA/EGMA results in Côte d'Ivoire.

Limited evidence of Long-Term Impact

Few studies have examined whether TaRL's effects are sustained over time. In the Balsakhi programme in India, positive effects faded significantly after one year. Overall effects were no longer statistically significant, and although gains among the lowest-performing students persisted, they were reduced in magnitude. In contrast, in Ghana effects persisted when children were re-tested one year after programme completion and in Columbia, the impact persisted by the end of grade 4 but with a 40 percent reduction.

Adding Instructional Time

In interpreting the effects of TaRL interventions, it is important to consider that in several cases, such as the programmes in Ghana, Madagascar, or the summer reading camps in Uttar Pradesh, TaRL was delivered outside regular instructional hours, effectively increasing the amount of time students spent on foundational learning. This raises a critical question: are the observed gains in learning attributable to the specific pedagogical features of the TaRL approach (e.g., assessment-led grouping, tailored instruction, interactive methods), or are they primarily a result of additional instructional time focused on basic skills? While both mechanisms may contribute, distinguishing between them is essential for assessing the true value-added of the approach and for making informed decisions about scalability, cost-effectiveness, and integration into regular school schedules. Future studies should aim to disentangle these effects by comparing TaRL-style instruction delivered during versus outside school hours or by using designs that hold time constant across treatment and control groups.

Multi-component interventions

Assessing TaRL's impact is further complicated when it is bundled with other interventions. In Colombia, the remedial programme emphasized phonics—unlike regular instruction, which combined whole-language and syllabic approaches—making it impossible to separate the effects of TaRL from those of a phonics-based approach. In Madagascar, TaRL was combined with a structured school management strengthening initiative, again making attribution difficult.

On the other hand, research in Botswana (Angrist and Meager 2023) illustrates how components can be isolated. There, A/B testing compared alternative methods of forming instructional groups within TaRL, helping to identify effective delivery models and refine programme design.

7 CRITIQUES OF THE TARL METHOD

TaRL has been subject to a range of critiques (Little 2020). These reflect both conceptual debates about its theory of change and practical concerns about implementation, sustainability, and generalizability. This section synthesises key critiques emerging from the literature, field experience, and evaluations to date.

Implementation, Sustainability, and Institutionalization

One of the most common critiques concerns the difficulty of achieving high-fidelity implementation, particularly within government systems (Banerjee 2010). Across contexts, studies consistently show that when TaRL is delivered by NGOs teams, well-trained facilitators, or external assistants, learning gains are higher. By contrast, when teachers in government schools are expected to implement TaRL independently, outcomes are more modest or mixed. These findings highlight that TaRL's effectiveness depends not only on its pedagogical design but also on the enabling environment in which it is implemented.

Closely linked to fidelity is the question of sustainability. Many of the strongest results have emerged from NGO-led or short-term pilot models, which benefit from dedicated staff, intensive monitoring, and focused support. However, scaling TaRL through government systems has proven more fragile: uptake often declines once external funding or technical assistance ends. Institutionalizing TaRL within national education systems raises unresolved questions about its role: Should it be integrated into the regular school day or delivered as additional instructional time? Is it best conceived as a temporary remedial measure or as a permanent fixture? How should it align with national curricula and assessment frameworks?

This ambiguity around institutionalization complicates planning, budgeting, and policymaking. Some countries have piloted government-led scale-ups, but there is little guidance or consensus on what sustainable adoption entails. Without clarity on whether TaRL is intended to remain a parallel system, be used intermittently as remediation, or evolve into a permanent instructional practice, governments risk inconsistent implementation and weak policy alignment.

Curriculum Misalignment and Policy Fit

Experts agree that remedial instruction requires strong system support (World Bank 2023). TaRL can conflict with national curricula and policy expectations, particularly where teachers are under pressure to "cover the syllabus." Because TaRL emphasizes teaching at the level of the child rather than the grade, it may be seen as diverging from mandated pacing guides or exam requirements. Without policy flexibility and alignment, teacher motivation and fidelity of implementation can be undermined.

Defining the Model and Evaluating Variants

As TaRL has spread globally, the term has become increasingly broad and loosely defined, making evaluation inconsistent. Some interventions with partial alignment—such as grouping without targeted instruction, or diagnostics without regrouping—are still labelled "TaRL." This blurring of definitions can inflate claims of impact or obscure which components actually drive results. For example, while both tracking and TaRL involve ability-based grouping, they differ substantially in structure and pedagogy. Without clear typologies and reporting standards, it remains difficult to identify which adaptations are effective and why.

Reintegration Challenges and the Nature of Remediation

TaRL is fundamentally remedial, targeting foundational literacy and numeracy that many students miss in early grades. Yet there is limited clarity on how students transition back into the standard curriculum once they acquire these foundational skills. In many contexts, no explicit "bridge" exists to help students reintegrate with grade-level instruction. This raises important questions: At what point are students expected to catch up? How can systems ensure they do not fall behind again? Without reintegration strategies, students risk being permanently left behind or isolated from regular instruction.

These critiques underscore a central tension in the TaRL model: while it offers a promising solution to foundational learning gaps, its success depends on implementation quality, system alignment, and long-term sustainability. As countries move from pilots to national strategies, these issues merit deeper attention.

8 CONCLUSION

This review has brought together the current evidence on TaRL, focusing on studies that meet basic standards of rigor. The findings show that TaRL can generate improvements in foundational literacy and numeracy, particularly for students who are furthest behind but its effectiveness is inconsistent and context-dependent.

The magnitude and consistency of impacts vary considerably depending on how the model is designed and delivered. The strongest results have been observed in short-term, intensive models implemented with close NGO supervision or by trained volunteers, while effects are weaker or absent in government-led programmes that lack dedicated time, sustained coaching, or sufficient resources. Gains are also concentrated in the most basic skills, with little evidence of progression into higher-order competencies, and the evidence on longer-term impacts remains limited.

As TaRL expands, questions of scalability, institutionalization, and alignment with national curricula become increasingly pressing. The diversity of models tested shows both the adaptability of the approach and the risks of dilution when key ingredients are missing. Critiques of TaRL emphasize that its success depends not only on pedagogical design but also on system-level factors such as teacher capacity, ongoing support, and policy fit. Taken together, the evidence suggests that TaRL can provide a valuable remedial strategy for boosting foundational skills, but its long-term role within national education systems remains uncertain. Ongoing evaluations in Zambia, Botswana, and Nigeria will be crucial for determining whether TaRL can be sustained and scaled within government systems while maintaining fidelity and impact.

REFERENCES

Akyeampong, K.; Andrabi, T.; Banerjee, A.; Banerji, R.; Dynarski, S.; Glennerster, R.; Grantham-McGregor, S.; Muralidharan, K.; Piper, B.; Ruto, S; Saavedra., J.; Schmelkes, S.; Yoshikawa, H. (2023). Cost-effective Approaches to Improve Global Learning: What does Recent Evidence Tell Us are Smart Buys for Improving Learning in Low- and Middle-income Countries? (English). Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/099420106132331608

Angrist, N. & Meager, R. (2023). Implementation Matters: Generalizing Treatment Effects in Education. https://edworkingpapers.com/sites/default/files/ai23-802.pdf

Alvarez Marinelli, H., Berlinski, S., & Busso, M. (2024). Remedial Education: Evidence from a Sequence of Experiments in Colombia. Journal of Human Resources 59(1), 141-174. https://muse.jhu.edu/article/917109.

Banerjee, A., Cole, S., Duflo, E. & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. Quarterly Journal of Economics, 122(3), 1235–1264. https://doi.org/10.1162/qjec.122.3.1235

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R. & Khemani, S. (2010). Pitfalls of participatory programmes: Evidence from a randomized evaluation in education in India. American Economic Journal: Applied Economics, 2(1), 1–30. https://doi.org/10.1257/pol.2.1.1

Banerjee, A., Banerji, R., Duflo, E. & Sethi, S. (2015). The impact of short-term remedial instruction on learning outcomes: Evidence from India. NBER Working Paper No. 20674.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M. & Walton, M. (2016). Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India. https://www.povertyactionlab.org/sites/default/files/researchpaper/TaRL_Paper_August2016.pdf

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M. & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions with an application. Journal of Economic Perspectives, 31(4), 73–102. https://doi.org/10.1257/jep.31.4.73

Bold, T., Filmer, D., Martin, G., Molina, E., Rockmore, C., Stacy, B., & Wane, W. (2019). What Do Teachers Know and Do? Does It Matter? Journal of Human Resources.

Coe, R. (2002) It's the Effect Size, Stupid: What effect size is and why it is important Paper presented at the British Educational Research Association annual conference, Exeter, 12- 14 September 2002

de Barros, A., Lubozha, T., Muraguri, P. and Vromant, N. (2023). A Randomized Evaluation of the Catch Up Programme in Zambia: Baseline Report. GPE KIX.

Duflo, E., Dupas, P. & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. American Economic Review, 101(5), 1739–1774. https://doi.org/10.1257/aer.101.5.1739

Duflo, E., Berry, J., Mukerji, S., & Shortland, M. (2015). A Wide Angle View of Learning: Evaluation of the CCE and LEP Programmemes in Haryana, India. 3ie Impact Evaluation Report 22. https://www.3ieimpact.org/sites/default/files/2017-11/ie_22_evaluation_of_cce_and_lep_in_haryana.pdf

Duflo, A., Kiessel, J., Lucas, A. & Shotland, M. (2020). Experimental evidence on four policies to increase learning at scale. NBER Working Paper No. 27298. https://www.nber.org/papers/w27298

Filmer, D., Rogers, F. H., Angrist, N., & Sabarwal, S. (2020). Learning-Adjusted Years of Schooling (LAYS): Defining a New Macro Measure of Education. Economics of Education Review.

Glewwe, P., Hanushek, E., Humpage, S., & Ravina, R. (2011). School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010. NBER Working Paper.

Global Education Evidence Advisory Panel. (2020). Cost-Effective Approaches to Improve Global Learning: What Does Recent Evidence Tell Us Are "Smart Buys" for Improving Learning in Low- and Middle-Income Countries? Washington, DC; London; New York: World Bank Group.

Global Education Evidence Advisory Panel. (2022).
Prioritizing learning during COVID-19: The most effective ways to keep children learning during and postpandemic.
[K. Akyeampong, T. Andrabi, A. Banerjee, R. Banerji, S. Dynarski, R. Glennerster, S. Grantham-McGregor, K. Muralidharan, B. Piper, S. Ruto, J. Saavedra, S. Schmelkes, H. Yoshikawa]. Washington D.C., London, Florence: The World Bank, FCDO, and UNICEF Office of Research - Innocenti.

Hanushek, E. A., & Woessmann, L. (2008).
The Role of Cognitive Skills in Economic Development. Journal of Economic Literature.

Hoadley, U. (2018).
Pedagogy in Poverty: Lessons from 20 years of curriculum reform in South Africa. London: Routledge.

Innovations for Poverty Action. (2017).
Catch Up Pilot using Teaching at the Right Level Methodology: Process Monitoring Report.
https://www.unicef.org/zambia/media/591/file/Zambia-catch-up-report.pdf

Lipovsek, V., Poswell, L., Morrell, A., Pershad, D., Vromant, N. and Grindle, A. Reflections on systems practices: implementing Teaching at the Right Level in Zambia, in Faul, M.V., Savage, L. and Edward Elgar Publishing, publisher (2023)
Systems Thinking in International Education and Development : Unlocking Learning for All? / edited by Moira V. Faul and Laura Savage.
https://www.elgaronline.com/downloadpdf/edcollchap-oa/book/9781802205930/book-part-9781802205930-12.pdf

Little, A. W. (2020).
Cost-effective approaches to improve global learning: Is Teaching at the Right Level (TaRL) a "good buy"? UKFIET blog.
https://www.ukfiet.org/2020/cost-effective-approaches-to-improve-global-learning-is-teaching-at-the-right-level-tarl-a-good-buy/
accessed on October 12, 2025

Maruyama, H. & Igei, Y. (2024).
Community-Wide Support for Primary Students to Improve Foundational Literacy and Numeracy: Empirical Evidence from Madagascar. Economic Development and Cultural Change 72 (4).
Available at: https://doi.org/10.1086/726178

Pratham. (n.d.).
Teaching at the Right Level. Retrieved April 14, 2025, from https://www.pratham.org/about/teaching-at-the-right-level/

Pritchett, L., & Beatty, A. (2015).
The Negative Consequences of Overambitious Curricula in Developing Countries.
CGD Working Paper.

Simpson, A. (2017).
The misdirection of public policy: Comparing and combining standardised effect sizes. Journal of Education Policy, 32(4), 450-466.

Stern, J. M. B., and Piper, B. (2019).
Resetting Targets: Why Large Effect Sizes in Education Development Programmes Are Not Resulting in More Students Reading at Benchmark.
RTI Press Publication No. OP-0060-1904. Research Triangle Park, NC: RTI Press.
https://doi. org/10.3768/rtipress.2019.op.0060.1904

TaRL Africa. (2025a). Classroom Methodology: An Evidence-Based Approach.
Retrieved April 14, 2025, from
https://teachingattherightlevel.org/classroom-methodology/

TaRL Africa. (2025b).
Botswana. Retrieved September 17, 2025, from
https://teachingattherightlevel.org/where-we-support/botswana/

TaRL Africa. (2025c).
Nigeria. Retrieved September 17, 2025, from
https://teachingattherightlevel.org/where-we-support/nigeria/

UNESCO (2021).
Global Education Monitoring Report 2021. Paris: UNESCO.

Wolf, S., Ogan, A., Jasińska, K., Kembou, S., Kamath Cannanure, V., Whitehead, H., Corbett, J., Tanoh, F., Wortsman, B., Lim, J., Zhao, H., & Kumar, A. (2025).
Impacts of targeted instruction with mobile-based support on teachers and students in rural Côte d'Ivoire.
https://www.researchgate.net/publication/382766652_Impacts_of_Targeted_Instruction_with_Mobile-based_Support_on_Teachers_and_Students_in_Rural_Cote_d'Ivoire

World Bank (2022).
The State of Global Learning Poverty: 2022 Update.

World Bank. (2023).
Learning recovery to acceleration: A global update on country efforts to improve learning and reduce inequalities. Washington, DC: World Bank.
https://doi.org/10.1596/978-1-4648-1965-3

APPENDIX

Location	Intervention	Evaluation Design	Sample Size	Assessment
Mumbai, Maharashtra	Balsakhi	RCT	Year 1(2): Grade 3 volunteer=32(38), Grade 4 volunteer=35(39) schools	Basic competencies taught in grades 1– 4
Vadodara, Gujarat			Year 1(2): Grade 3 volunteer=49(61), Grade 3 volunteer=49(61) schools	
Uttar Pradesh	reading camps	RCT	C=85, T=65 villages	ASER
Bihar	summer camps	RCT	C=40, T=120 villages	ASER
	materials only	RCT	C=39, T=37 villages	
	training		C=39, T=37 villages	
	training + volunteers		C=39, T=37 villages	
Uttarakhand	training	RCT	C=35, T=36 villages	ASER
	training + volunteers		C=35, T=39 villages	
Haryana	dedicated hour	RCT	C=100, T=100, Non-TaRL intervention=100, Combined=100 schools	ASER
Uttar Pradesh	materials only	RCT	C=123, T=119 schools	ASER
	four 10-day camps		C=123, T=122 schools	
	two 20-day camps		C=123, T=120 schools	
Assam	study group	RCT	C=100, T=100 villages	ASER
	three 10-day camps		C=100, T=100 villages	
	study group + camp		C=100, T=100 villages	
Kenya	tracking	RCT	C=61, T=60	Math (counting to subtracting 3-digit numbers), literacy (identifying letters to reading and understanding sentences)
Ghana	remedial pull-out	RCT	C=100, T=100	Bespoke exam covering parts of curriculum and including foundational literacy and numeracy similar to ASER.
	remedial after-school		C=100, T=100	
	partial day tracking		C=100, T=100	
Madagascar	TaRL plus school management	RCT	C=70, T=70	ASER reading, 44-item written math (number recognition, addition, subtraction, multiplication, division, and word problems)
Colombia	remedial education	RCT	Cohort 1: C=49, T=49 Cohort 2: C=42, T=42 Cohort 3: C=40, T=40	EGRA (letter-sounds, pseudoword-reading, oral reading fluency, comprehension)
Côte d'Ivoire	TaRL	RCT	C=85, T=82	ASER, EGRA (letter-reading, word-reading, and pseudoword-reading), EGMA (missing value, addition, and subtraction)

AFLEARN
DATAHUB

DataFirst

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

African Foundational Learning Data Hub
DataFirst
University of Cape Town
www.datafirst.uct.ac.za/aflearn